





CDS	complement (27277..27552) /note="unnamed protein product; gene_id:MPE11.7 /codon_start=1 /evidence=not_experimental /protein_id="BAB01053.1" /db_xref="GI:9279595" /translation="MTHAREMRSSLTITLLVILLSYMLHLFCVTSRVGAIRIPETP ASGRQEDLMKYFGAGFPVPFVGVGIGESKRIVSPSCDPLHN" 30765..31883 /note="unnamed protein product; gb AAF16548.1 gene_id:MPE11.8 similar to unknown protein" /codon_start=1 /evidence=not_experimental /protein_id="BAB01054.1" /db_xref="GI:9279596" /translation="MPKERSKVSLLDKYRSPLCCEASLALPKSEKQVKEWEARCP VMEHPHNGILLICSYNGCRPYMCDTSHRHSNCPDQPRKASKKPKSLRLBEEES NEPTMEDVSDSTAVNLGEAASEITVVDLSDGERGEVEEVEEVEEVEEVEEIV TTEEDQKNPKOKLTCPLCRGHKEWVVVKAARCFWNSKHSRSCSETCFPSGYSGLR KIALHLPVVRSEADPERQSRWRRLERQSDGLDLSLTSQSFSGDEISNDGFLPAD TKLLTVFLVFRPSSSSSSSGTSGRKRHTSGRRSRKPSASLWGESYEGNTGT SPRDENNQSDQVSGTRRRSRRTVTDDEDEEP" complement (join (32197..32430,32525..32676,32969..33332, 33404..33487)) /note="gene_id:MPE11.9" /codon_start=1 /evidence=not_experimental /product="chloroplast 50S ribosomal protein L15" /protein_id="BAB01055.1" /db_xref="GI:9279597" /translation="MATPLSISNPLTSBHCYRLHLSSTSFKNVSVLGNPQSLILSL KINQTLNRQQAFPLVVSQTAATSAVVAAPERLNDLNGPQGRKKQKRGKRG ISAGQACGMRGQKSGIMRGFGQYALYRLPKLGRAGMRSGLPKLYLP VNIKDTAGQEGDEVSLETIKQGLINPSGRERKLPKILGTSLSKLFTKARAF STQAKCLEASGCTLVLPGRKKWVPSVAKNQARADEYFAKRAAAAEATSEPAAS A" complement (join (33961..34021,34114..34226,34391..34482, 34602..34800)) /note="unnamed protein product; gb AAF26483.1 gene_id:MPE11.10 similar to unknown protein" /codon_start=1 /evidence=not_experimental /protein_id="BAB01056.1" /db_xref="GI:9279598" /translation="MKNVLMII DESNASYDLIIWALENOKDTIESSKVYIFAKQPONS FTPTVLSSVGAQIYPFSPNSLIRLAQENKMKIALGILEKAKKICLNHGIIKAEI FTDGPDKILRIQERNILNIVTSQSLKKTONTDCSLVVKLRKD" join (36033..36150,36243..36283,36497..36565,36651..36707, 36797..36871) /note="unnamed protein product; contains similarity to transcription-associated zinc ribbon protein gene_id:MPE11.11" /codon_start=1 /evidence=not_experimental /protein_id="BAB01057.1" /db_xref="GI:9279599" /translation="MEKRSSEFTFNCLGTMVLVKSTKYAECPHCKTTRNAKDIIDK EIAVTSARDIRELIGISLFGENTQAEALPKIKACEKQHPVELVYVTRQTSRDEG QTYTTCPCNAHRFTSG" join (37285..37965,38579..38674)	Db	12043	ATAAAGTGTGACGTCATGATCTTACGCTTTTAAAAACATCGCATGATGTCAATTAGCAT	11984
	/evidence=not_experimental /protein_id="BAB01053.1" /db_xref="GI:9279595" /translation="MTHAREMRSSLTITLLVILLSYMLHLFCVTSRVGAIRIPETP ASGRQEDLMKYFGAGFPVPFVGVGIGESKRIVSPSCDPLHN" 30765..31883 /note="unnamed protein product; gb AAF16548.1 gene_id:MPE11.8 similar to unknown protein" /codon_start=1 /evidence=not_experimental /protein_id="BAB01054.1" /db_xref="GI:9279596" /translation="MPKERSKVSLLDKYRSPLCCEASLALPKSEKQVKEWEARCP VMEHPHNGILLICSYNGCRPYMCDTSHRHSNCPDQPRKASKKPKSLRLBEEES NEPTMEDVSDSTAVNLGEAASEITVVDLSDGERGEVEEVEEVEEVEEVEEIV TTEEDQKNPKOKLTCPLCRGHKEWVVVKAARCFWNSKHSRSCSETCFPSGYSGLR KIALHLPVVRSEADPERQSRWRRLERQSDGLDLSLTSQSFSGDEISNDGFLPAD TKLLTVFLVFRPSSSSSSSGTSGRKRHTSGRRSRKPSASLWGESYEGNTGT SPRDENNQSDQVSGTRRRSRRTVTDDEDEEP" complement (join (32197..32430,32525..32676,32969..33332, 33404..33487)) /note="gene_id:MPE11.9" /codon_start=1 /evidence=not_experimental /product="chloroplast 50S ribosomal protein L15" /protein_id="BAB01055.1" /db_xref="GI:9279597" /translation="MATPLSISNPLTSBHCYRLHLSSTSFKNVSVLGNPQSLILSL KINQTLNRQQAFPLVVSQTAATSAVVAAPERLNDLNGPQGRKKQKRGKRG ISAGQACGMRGQKSGIMRGFGQYALYRLPKLGRAGMRSGLPKLYLP VNIKDTAGQEGDEVSLETIKQGLINPSGRERKLPKILGTSLSKLFTKARAF STQAKCLEASGCTLVLPGRKKWVPSVAKNQARADEYFAKRAAAAEATSEPAAS A" complement (join (33961..34021,34114..34226,34391..34482, 34602..34800)) /note="unnamed protein product; gb AAF26483.1 gene_id:MPE11.10 similar to unknown protein" /codon_start=1 /evidence=not_experimental /protein_id="BAB01056.1" /db_xref="GI:9279598" /translation="MKNVLMII DESNASYDLIIWALENOKDTIESSKVYIFAKQPONS FTPTVLSSVGAQIYPFSPNSLIRLAQENKMKIALGILEKAKKICLNHGIIKAEI FTDGPDKILRIQERNILNIVTSQSLKKTONTDCSLVVKLRKD" join (36033..36150,36243..36283,36497..36565,36651..36707, 36797..36871) /note="unnamed protein product; contains similarity to transcription-associated zinc ribbon protein gene_id:MPE11.11" /codon_start=1 /evidence=not_experimental /protein_id="BAB01057.1" /db_xref="GI:9279599" /translation="MEKRSSEFTFNCLGTMVLVKSTKYAECPHCKTTRNAKDIIDK EIAVTSARDIRELIGISLFGENTQAEALPKIKACEKQHPVELVYVTRQTSRDEG QTYTTCPCNAHRFTSG" join (37285..37965,38579..38674)	Qy	121	CAATCTCCACCGTCCAAATTTTATTAGTTTGTGACAAATATCGACCGTCTTAAGTTCCACACC	180
	/evidence=not_experimental /protein_id="BAB01053.1" /db_xref="GI:9279595" /translation="MTHAREMRSSLTITLLVILLSYMLHLFCVTSRVGAIRIPETP ASGRQEDLMKYFGAGFPVPFVGVGIGESKRIVSPSCDPLHN" 30765..31883 /note="unnamed protein product; gb AAF16548.1 gene_id:MPE11.8 similar to unknown protein" /codon_start=1 /evidence=not_experimental /protein_id="BAB01054.1" /db_xref="GI:9279596" /translation="MPKERSKVSLLDKYRSPLCCEASLALPKSEKQVKEWEARCP VMEHPHNGILLICSYNGCRPYMCDTSHRHSNCPDQPRKASKKPKSLRLBEEES NEPTMEDVSDSTAVNLGEAASEITVVDLSDGERGEVEEVEEVEEVEEVEEIV TTEEDQKNPKOKLTCPLCRGHKEWVVVKAARCFWNSKHSRSCSETCFPSGYSGLR KIALHLPVVRSEADPERQSRWRRLERQSDGLDLSLTSQSFSGDEISNDGFLPAD TKLLTVFLVFRPSSSSSSSGTSGRKRHTSGRRSRKPSASLWGESYEGNTGT SPRDENNQSDQVSGTRRRSRRTVTDDEDEEP" complement (join (32197..32430,32525..32676,32969..33332, 33404..33487)) /note="gene_id:MPE11.9" /codon_start=1 /evidence=not_experimental /product="chloroplast 50S ribosomal protein L15" /protein_id="BAB01055.1" /db_xref="GI:9279597" /translation="MATPLSISNPLTSBHCYRLHLSSTSFKNVSVLGNPQSLILSL KINQTLNRQQAFPLVVSQTAATSAVVAAPERLNDLNGPQGRKKQKRGKRG ISAGQACGMRGQKSGIMRGFGQYALYRLPKLGRAGMRSGLPKLYLP VNIKDTAGQEGDEVSLETIKQGLINPSGRERKLPKILGTSLSKLFTKARAF STQAKCLEASGCTLVLPGRKKWVPSVAKNQARADEYFAKRAAAAEATSEPAAS A" complement (join (33961..34021,34114..34226,34391..34482, 34602..34800)) /note="unnamed protein product; gb AAF26483.1 gene_id:MPE11.10 similar to unknown protein" /codon_start=1 /evidence=not_experimental /protein_id="BAB01056.1" /db_xref="GI:9279598" /translation="MKNVLMII DESNASYDLIIWALENOKDTIESSKVYIFAKQPONS FTPTVLSSVGAQIYPFSPNSLIRLAQENKMKIALGILEKAKKICLNHGIIKAEI FTDGPDKILRIQERNILNIVTSQSLKKTONTDCSLVVKLRKD" join (36033..36150,36243..36283,36497..36565,36651..36707, 36797..36871) /note="unnamed protein product; contains similarity to transcription-associated zinc ribbon protein gene_id:MPE11.11" /codon_start=1 /evidence=not_experimental /protein_id="BAB01057.1" /db_xref="GI:9279599" /translation="MEKRSSEFTFNCLGTMVLVKSTKYAECPHCKTTRNAKDIIDK EIAVTSARDIRELIGISLFGENTQAEALPKIKACEKQHPVELVYVTRQTSRDEG QTYTTCPCNAHRFTSG" join (37285..37965,38579..38674)	Db	11983	CAATCTCCACCGTCCAAATTTTATTAGTTTGTGACAAATATCGACCGTCTTAAGTTCCACACC	11924
CDS					
	/evidence=not_experimental /protein_id="BAB01053.1" /db_xref="GI:9279595" /translation="MTHAREMRSSLTITLLVILLSYMLHLFCVTSRVGAIRIPETP ASGRQEDLMKYFGAGFPVPFVGVGIGESKRIVSPSCDPLHN" 30765..31883 /note="unnamed protein product; gb AAF16548.1 gene_id:MPE11.8 similar to unknown protein" /codon_start=1 /evidence=not_experimental /protein_id="BAB01054.1" /db_xref="GI:9279596" /translation="MPKERSKVSLLDKYRSPLCCEASLALPKSEKQVKEWEARCP VMEHPHNGILLICSYNGCRPYMCDTSHRHSNCPDQPRKASKKPKSLRLBEEES NEPTMEDVSDSTAVNLGEAASEITVVDLSDGERGEVEEVEEVEEVEEVEEIV TTEEDQKNPKOKLTCPLCRGHKEWVVVKAARCFWNSKHSRSCSETCFPSGYSGLR KIALHLPVVRSEADPERQSRWRRLERQSDGLDLSLTSQSFSGDEISNDGFLPAD TKLLTVFLVFRPSSSSSSSGTSGRKRHTSGRRSRKPSASLWGESYEGNTGT SPRDENNQSDQVSGTRRRSRRTVTDDEDEEP" complement (join (32197..32430,32525..32676,32969..33332, 33404..33487)) /note="gene_id:MPE11.9" /codon_start=1 /evidence=not_experimental /product="chloroplast 50S ribosomal protein L15" /protein_id="BAB01055.1" /db_xref="GI:9279597" /translation="MATPLSISNPLTSBHCYRLHLSSTSFKNVSVLGNPQSLILSL KINQTLNRQQAFPLVVSQTAATSAVVAAPERLNDLNGPQGRKKQKRGKRG ISAGQACGMRGQKSGIMRGFGQYALYRLPKLGRAGMRSGLPKLYLP VNIKDTAGQEGDEVSLETIKQGLINPSGRERKLPKILGTSLSKLFTKARAF STQAKCLEASGCTLVLPGRKKWVPSVAKNQARADEYFAKRAAAAEATSEPAAS A" complement (join (33961..34021,34114..34226,34391..34482, 34602..34800)) /note="unnamed protein product; gb AAF26483.1 gene_id:MPE11.10 similar to unknown protein" /codon_start=1 /evidence=not_experimental /protein_id="BAB01056.1" /db_xref="GI:9279598" /translation="MKNVLMII DESNASYDLIIWALENOKDTIESSKVYIFAKQPONS FTPTVLSSVGAQIYPFSPNSLIRLAQENKMKIALGILEKAKKICLNHGIIKAEI FTDGPDKILRIQERNILNIVTSQSLKKTONTDCSLVVKLRKD" join (36033..36150,36243..36283,36497..36565,36651..36707, 36797..36871) /note="unnamed protein product; contains similarity to transcription-associated zinc ribbon protein gene_id:MPE11.11" /codon_start=1 /evidence=not_experimental /protein_id="BAB01057.1" /db_xref="GI:9279599" /translation="MEKRSSEFTFNCLGTMVLVKSTKYAECPHCKTTRNAKDIIDK EIAVTSARDIRELIGISLFGENTQAEALPKIKACEKQHPVELVYVTRQTSRDEG QTYTTCPCNAHRFTSG" join (37285..37965,38579..38674)	Qy	181	GACGGCTATAAGAGTTTCATTATAAATTTTACGAAAAATAAATACGAAAAATAATTTTTC	240
	/evidence=not_experimental /protein_id="BAB01053.1" /db_xref="GI:9279595" /translation="MTHAREMRSSLTITLLVILLSYMLHLFCVTSRVGAIRIPETP ASGRQEDLMKYFGAGFPVPFVGVGIGESKRIVSPSCDPLHN" 30765..31883 /note="unnamed protein product; gb AAF16548.1 gene_id:MPE11.8 similar to unknown protein" /codon_start=1 /evidence=not_experimental /protein_id="BAB01054.1" /db_xref="GI:9279596" /translation="MPKERSKVSLLDKYRSPLCCEASLALPKSEKQVKEWEARCP VMEHPHNGILLICSYNGCRPYMCDTSHRHSNCPDQPRKASKKPKSLRLBEEES NEPTMEDVSDSTAVNLGEAASEITVVDLSDGERGEVEEVEEVEEVEEVEEIV TTEEDQKNPKOKLTCPLCRGHKEWVVVKAARCFWNSKHSRSCSETCFPSGYSGLR KIALHLPVVRSEADPERQSRWRRLERQSDGLDLSLTSQSFSGDEISNDGFLPAD TKLLTVFLVFRPSSSSSSSGTSGRKRHTSGRRSRKPSASLWGESYEGNTGT SPRDENNQSDQVSGTRRRSRRTVTDDEDEEP" complement (join (32197..32430,32525..32676,32969..33332, 33404..33487)) /note="gene_id:MPE11.9" /codon_start=1 /evidence=not_experimental /product="chloroplast 50S ribosomal protein L15" /protein_id="BAB01055.1" /db_xref="GI:9279597" /translation="MATPLSISNPLTSBHCYRLHLSSTSFKNVSVLGNPQSLILSL KINQTLNRQQAFPLVVSQTAATSAVVAAPERLNDLNGPQGRKKQKRGKRG ISAGQACGMRGQKSGIMRGFGQYALYRLPKLGRAGMRSGLPKLYLP VNIKDTAGQEGDEVSLETIKQGLINPSGRERKLPKILGTSLSKLFTKARAF STQAKCLEASGCTLVLPGRKKWVPSVAKNQARADEYFAKRAAAAEATSEPAAS A" complement (join (33961..34021,34114..34226,34391..34482, 34602..34800)) /note="unnamed protein product; gb AAF26483.1 gene_id:MPE11.10 similar to unknown protein" /codon_start=1 /evidence=not_experimental /protein_id="BAB01056.1" /db_xref="GI:9279598" /translation="MKNVLMII DESNASYDLIIWALENOKDTIESSKVYIFAKQPONS FTPTVLSSVGAQIYPFSPNSLIRLAQENKMKIALGILEKAKKICLNHGIIKAEI FTDGPDKILRIQERNILNIVTSQSLKKTONTDCSLVVKLRKD" join (36033..36150,36243..36283,36497..36565,36651..36707, 36797..36871) /note="unnamed protein product; contains similarity to transcription-associated zinc ribbon protein gene_id:MPE11.11" /codon_start=1 /evidence=not_experimental /protein_id="BAB01057.1" /db_xref="GI:9279599" /translation="MEKRSSEFTFNCLGTMVLVKSTKYAECPHCKTTRNAKDIIDK EIAVTSARDIRELIGISLFGENTQAEALPKIKACEKQHPVELVYVTRQTSRDEG QTYTTCPCNAHRFTSG" join (37285..37965,38579..38674)	Db	11923	GACGGCTATAAGAGTTTCATTATAAATTTTACGAAAAATAAATACGAAAAATAATTTTTC	11864
	/evidence=not_experimental /protein_id="BAB01053.1" /db_xref="GI:9279595" /translation="MTHAREMRSSLTITLLVILLSYMLHLFCVTSRVGAIRIPETP ASGRQEDLMKYFGAGFPVPFVGVGIGESKRIVSPSCDPLHN" 30765..31883 /note="unnamed protein product; gb AAF16548.1 gene_id:MPE11.8 similar to unknown protein" /codon_start=1 /evidence=not_experimental /protein_id="BAB01054.1" /db_xref="GI:9279596" /translation="MPKERSKVSLLDKYRSPLCCEASLALPKSEKQVKEWEARCP VMEHPHNGILLICSYNGCRPYMCDTSHRHSNCPDQPRKASKKPKSLRLBEEES NEPTMEDVSDSTAVNLGEAASEITVVDLSDGERGEVEEVEEVEEVEEVEEIV TTEEDQKNPKOKLTCPLCRGHKEWVVVKAARCFWNSKHSRSCSETCFPSGYSGLR KIALHLPVVRSEADPERQSRWRRLERQSDGLDLSLTSQSFSGDEISNDGFLPAD TKLLTVFLVFRPSSSSSSSGTSGRKRHTSGRRSRKPSASLWGESYEGNTGT SPRDENNQSDQVSGTRRRSRRTVTDDEDEEP" complement (join (32197..32430,32525..32676,32969..33332, 33404..33487)) /note="gene_id:MPE11.9" /codon_start=1 /evidence=not_experimental /product="chloroplast 50S ribosomal protein L15" /protein_id="BAB01055.1" /db_xref="GI:9279597" /translation="MATPLSISNPLTSBHCYRLHLSSTSFKNVSVLGNPQSLILSL KINQTLNRQQAFPLVVSQTAATSAVVAAPERLNDLNGPQGRKKQKRGKRG ISAGQACGMRGQKSGIMRGFGQYALYRLPKLGRAGMRSGLPKLYLP VNIKDTAGQEGDEVSLETIKQGLINPSGRERKLPKILGTSLSKLFTKARAF STQAKCLEASGCTLVLPGRKKWVPSVAKNQARADEYFAKRAAAAEATSEPAAS A" complement (join (33961..34021,34114..34226,34391..34482, 34602..34800)) /note="unnamed protein product; gb AAF26483.1 gene_id:MPE11.10 similar to unknown protein" /codon_start=1 /evidence=not_experimental /protein_id="BAB01056.1" /db_xref="GI:9279598" /translation="MKNVLMII DESNASYDLIIWALENOKDTIESSKVYIFAKQPONS FTPTVLSSVGAQIYPFSPNSLIRLAQENKMKIALGILEKAKKICLNHGIIKAEI FTDGPDKILRIQERNILNIVTSQSLKKTONTDCSLVVKLRKD" join (36033..36150,36243..36283,36497..36565,36651..36707, 36797..36871) /note="unnamed protein product; contains similarity to transcription-associated zinc ribbon protein gene_id:MPE11.11" /codon_start=1 /evidence=not_experimental /protein_id="BAB01057.1" /db_xref="GI:9279599" /translation="MEKRSSEFTFNCLGTMVLVKSTKYAECPHCKTTRNAKDIIDK EIAVTSARDIRELIGISLFGENTQAEALPKIKACEKQHPVELVYVTRQTSRDEG QTYTTCPCNAHRFTSG" join (37285..37965,38579..38674)	Qy	241	TTGACTAAGCTTAAACGACGCGTTAAACATTTTC	274
	/evidence=not_experimental /protein_id="BAB01053.1" /db_xref="GI:9279595" /translation="MTHAREMRSSLTITLLVILLSYMLHLFCVTSRVGAIRIPETP ASGRQEDLMKYFGAGFPVPFVGVGIGESKRIVSPSCDPLHN" 30765..31883 /note="unnamed protein product; gb AAF16548.1 gene_id:MPE11.8 similar to unknown protein" /codon_start=1 /evidence=not_experimental /protein_id="BAB01054.1" /db_xref="GI:9279596" /translation="MPKERSKVSLLDKYRSPLCCEASLALPKSEKQVKEWEARCP VMEHPHNGILLICSYNGCRPYMCDTSHRHSNCPDQPRKASKKPKSLRLBEEES NEPTMEDVSDSTAVNLGEAASEITVVDLSDGERGEVEEVEEVEEVEEVEEIV TTEEDQKNPKOKLTCPLCRGHKEWVVVKAARCFWNSKHSRSCSETCFPSGYSGLR KIALHLPVVRSEADPERQSRWRRLERQSDGLDLSLTSQSFSGDEISNDGFLPAD TKLLTVFLVFRPSSSSSSSGTSGRKRHTSGRRSRKPSASLWGESYEGNTGT SPRDENNQSDQVSGTRRRSRRTVTDDEDEEP" complement (join (32197..32430,32525..32676,32969..33332, 33404..33487)) /note="gene_id:MPE11.9" /codon_start=1 /evidence=not_experimental /product="chloroplast 50S ribosomal protein L15" /protein_id="BAB01055.1" /db_xref="GI:9279597" /translation="MATPLSISNPLTSBHCYRLHLSSTSFKNVSVLGNPQSLILSL KINQTLNRQQAFPLVVSQTAATSAVVAAPERLNDLNGPQGRKKQKRGKRG ISAGQACGMRGQKSGIMRGFGQYALYRLPKLGRAGMRSGLPKLYLP VNIKDTAGQEGDEVSLETIKQGLINPSGRERKLPKILGTSLSKLFTKARAF STQAKCLEASGCTLVLPGRKKWVPSVAKNQARADEYFAKRAAAAEATSEPAAS A" complement (join (33961..34021,34114..34226,34391..34482, 34602..34800)) /note="unnamed protein product; gb AAF26483.1 gene_id:MPE11.10 similar to unknown protein" /codon_start=1 /evidence=not_experimental /protein_id="BAB01056.1" /db_xref="GI:9279598" /translation="MKNVLMII DESNASYDLIIWALENOKDTIESSKVYIFAKQPONS FTPTVLSSVGAQIYPFSPNSLIRLAQENKMKIALGILEKAKKICLNHGIIKAEI FTDGPDKILRIQERNILNIVTSQSLKKTONTDCSLVVKLRKD" join (36033..36150,36243..36283,36497..36565,36651..36707, 36797..36871) /note="unnamed protein product; contains similarity to transcription-associated zinc ribbon protein gene_id:MPE11.11" /codon_start=1 /evidence=not_experimental /protein_id="BAB01057.1" /db_xref="GI:9279599" /translation="MEKRSSEFTFNCLGTMVLVKSTKYAECPHCKTTRNAKDIIDK EIAVTSARDIRELIGISLFGENTQAEALPKIKACEKQHPVELVYVTRQTSRDEG QTYTTCPCNAHRFTSG" join (37285..37965,38579..38674)	Db	11863	TTGACTAAGCTTAAACGACGCGTTAAACATTTTC	11830
RESULT 3					
CS007929					
LOCUS	CS007929	1226 bp	DNA	linear	
DEFINITION	Sequence 2 from Patent EP1502953.			PAT 11-FEB-2005	
ACCESSION	CS007929				
VERSION	CS007929.1	GI:59668860			
KEYWORDS	Arabidopsis thaliana (thale cress)				
SOURCE	Arabidopsis thaliana				
ORGANISM	Arabidopsis thaliana				
REFERENCE	1				
AUTHORS	Pfitzner, A. and Roth, B.				
TITLE	Chemically-inducible promoters for the expression of proteins in plants				
JOURNAL	Patent: EP 1502953-A 2 02-FEB-2005;				
FEATURES	Location/Qualifiers				
source	1..1226				
	/organism="Arabidopsis thaliana"				
	/mol_type="unassigned DNA"				
	/db_xref="taxon:3702"				
ORIGIN					
Query Match	94.3%;	Score 258.4;	DB 6;	Length 1226;	
Best Local Similarity	99.6%;	Pred. No. 2.2e-48;			
Matches	259;	Conservative 0;	Mismatches 1;	Indels 0; Gaps 0;	
Qy	1	TCTAGAAATATAGCGATACCAATTTTCCACATGGACTTCCTTTATTTCCAAAAGTCA	60		
Db	967	TCTAGAAATATAGCGATACCAATTTTCCACATGGACTTCCTTTATTTCCAAAAGTCA	1026		
Qy	61	ATAAAGTGTGACGTCATGATCTTACGCTTTTAAAAACATCGCATGATGTCTATTAGCAT	120		
Db	1027	ATAAAGTGTGACGTCATGATCTTACGCTTTTAAAAACATCGCATGATGTCTATTAGCAT	1086		
Qy	121	CAATCTCCACCGTCCAAATTTTATTAGTTTGTGACAAATATCGACCGTCTTAAGTTCCACACC	180		
Db	1087	CAATCTCCACCGTCCAAATTTTATTAGTTTGTGACAAATATCGACCGTCTTAAGTTCCACACC	1146		
Qy	181	GACGGCTATAAGAGTTTCATTATAAATTTTACGAAAAATAAATACGAAAAATAATTTTTC	240		
Db	1147	GACGGCTATAAGAGTTTCATTATAAATTTTACGAAAAATAAATACGAAAAATAATTTTTC	1206		
Qy	241	TTGACTAAGCTTAAACGACG 260			
Db	1207	TTGACTAAGCTTAAACGACG 1226			
RESULT 4					
CS025770					
LOCUS	CS025770	1226 bp	DNA	linear	
DEFINITION	Sequence 2 from Patent WO2005014829.			PAT 03-MAR-2005	
ACCESSION	CS025770				
VERSION	CS025770.1	GI:60496427			
KEYWORDS					

SOURCE Arabidopsis thaliana (thale cress)  
 ORGANISM Arabidopsis thaliana  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;  
 rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsi  
 1  
 Pfitzner, A. and Roth, B.  
 Chemically-inducible promoters for the expression of proteins in  
 plants  
 Patent: WO 2005014829-A 2 17-FEB-2005;  
 Universitaet Hohenheim (DE)  
 JOURNAL Location/Qualifiers  
 FEATURES  
 source  
 1..1226  
 /organism="Arabidopsis thaliana"  
 /mol\_type="unassigned DNA"  
 /db\_xref="taxon:3702"  
 ORIGIN  
 Query Match 94.3%; Score 258.4; DB 6; Length 1226;  
 Best Local Similarity 99.4%; Pred. No. 2.2e-48;  
 Matches 259; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 QY 1 TCTAGAAATATAGCCGATACCAATTTTCCACATGGACTTCCTTTATTCCAAAGTCA 60  
 Db 967 TCTAAATATATAGCGATACCAATTTTCCACATGGACTTCCTTTATTCCAAAGTCA 1026  
 QY 61 ATAAAGTGTGAGTCATGATCTTACGCTTTTAAACATCGCATGATGTGATTAGCAT 120  
 Db 1027 ATAAAGTGTGAGTCATGATCTTACGCTTTTAAACATCGCATGATGTGATTAGCAT 1086  
 QY 121 CAATCTCCACCGTCCAAATTTATTAGTTGTTGACAAATATCGACCGCTTAAGTCCACACC 180  
 Db 1087 CAATCTCCACCGTCCAAATTTATTAGTTGTTGACAAATATCGACCGCTTAAGTCCACACC 1146  
 QY 181 GACGGCTATAAGAGTTTCATTAATAATTTTACCAAAATAAAATCAGCAATAATTTTTC 240  
 Db 1147 GACGGCTATAAGAGTTTCATTAATAATTTTACCAAAATAAAATCAGCAATAATTTTTC 1206  
 QY 241 TTGACTTAAGCTTAAACGACG 260  
 Db 1207 TTGACTTAAGCTTAAACGACG 1226  
 RESULT 5  
 AR488146  
 LOCUS  
 DEFINITION Sequence 1 from patent US 6706952.  
 ACCESSION AR488146  
 VERSION AR488146.1 GI:47253920  
 KEYWORDS  
 SOURCE Unknown.  
 ORGANISM Unclassified.  
 1 (bases 1 to 509)  
 Cad.R.M. and Dietrich, R.A.  
 Arabidopsis gene encoding a protein involved in the regulation of  
 SAR gene expression in plants  
 Patent: US 6706952-A 1 16-MAR-2004;  
 Syngenta Participations AG; Basel;  
 CHX;  
 FEATURES  
 source  
 1..509  
 /organism="unknown"  
 /mol\_type="genomic DNA"  
 ORIGIN  
 Query Match 19.3%; Score 52.8; DB 6; Length 509;  
 Best Local Similarity 96.4%; Pred. No. 0.085;  
 Matches 54; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
 QY 219 AAAATCAGCAATAAATTTTCTTGACTAAGCTTAAACGACGCGTTAAACATTTC 274  
 Db 1 AAAATCAGCAATAAATTTTCTTGACTAAGCTTAAACGACGCGTTAAACATTTC 56

RESULT 6  
 AKI75239  
 LOCUS  
 DEFINITION Arabidopsis thaliana mRNA for NIMIN-2, complete cds, clone:  
 RAFL21-66-M11.  
 AKI75239 544 bp mRNA linear PLN 09-SEP-2004  
 Arabidopsis thaliana mRNA for NIMIN-2, complete cds, clone:  
 RAFL21-66-M11.  
 AKI75239 GI:51968619  
 VERSION  
 KEYWORDS  
 SOURCE  
 ORGANISM  
 Arabidopsis thaliana (thale cress)  
 Arabidopsis thaliana  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;  
 rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsi  
 1  
 Totoki, Y., Seki, M., Ishida, J., Nakajima, M., Enju, A., Kamiya, A.,  
 Narusaka, M., Shin-i, T., Nakagawa, M., Sakamoto, N., Oishi, K.,  
 Kohara, Y., Kobayashi, M., Toyoda, A., Sakaki, Y., Sakurai, T., Iida, K.,  
 Akiyama, K., Satou, M., Toyoda, T., Konagaya, A., Carninci, P.,  
 Kawai, J., Hayashizaki, Y. and Shinozaki, K.  
 Large-scale analysis of RIKEN Arabidopsis full-length (RAFL) cDNAs  
 Unpublished  
 2 (bases 1 to 544)  
 Totoki, Y., Seki, M., Ishida, J., Nakajima, M., Enju, A., Kamiya, A.,  
 Narusaka, M., Shin-i, T., Nakagawa, M., Sakamoto, N., Oishi, K.,  
 Kohara, Y., Kobayashi, M., Toyoda, A., Sakaki, Y., Sakurai, T., Iida, K.,  
 Akiyama, K., Satou, M., Toyoda, T., Konagaya, A., Carninci, P.,  
 Kawai, J., Hayashizaki, Y. and Shinozaki, K.  
 Direct Submission  
 Submitted (06-SEP-2004) Motoaki Seki, RIKEN Genomic Sciences  
 Center; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa  
 230-0045, Japan (E-mail:msaki@gs.riken.jp,  
 URL:http://rage.gsc.riken.jp/, Tel:81-45-503-9625,  
 Fax:81-45-503-9586)  
 An Arabidopsis full-length cDNA library was constructed essentially  
 as reported previously (Seki et al. (1998) Plant J. 15:707-720;  
 Seki et al. (2002) Science 296:141-145). cDNA cleaved with BamHI  
 and XhoI was ligated to modified Lambda phage-1-E vector (Carninci et  
 al. (2001) Genomics 77:79-90) digested with BamHI and SalI. This  
 clone is in a modified pBluescript vector.  
 Please visit our web site (http://rage.gsc.riken.jp/) for further  
 details.  
 FEATURES  
 source  
 1..544  
 /organism="Arabidopsis thaliana"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:3702"  
 /chromosome="3"  
 /clone="RAFL21-66-M11"  
 /ecotype="Columbia"  
 /note="common name: thale cress"  
 gene  
 1..544  
 /gene="At3g25882"  
 CDS  
 62..430  
 /gene="At3g25882"  
 /codon\_start=1  
 /product="NIMIN-2"  
 /protein\_id="BAD43002.1"  
 /db\_xref="GI:51968620"  
 /translation="MNSIKKEERVEEDNGKSDGNRGKPSFTEVTVTEEVDEPFKIL  
 LRVHVATRTVAKVGVAGELPSKKRSQNLGLRSLDCNGYRDGFDEINRVGL  
 QQLGLDLNCKTEPDSVSLSL"

Query Match 18.2%; Score 50; DB 15; Length 544;  
 Best Local Similarity 100.0%; Pred. No. 0.36;  
 Matches 50; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 225 AGCAATAATTTTTTCTTGACTAAGCTTAAACGACGCGTTAAACATTTC 274  
 Db 1 AGCAATAATTTTTTCTTGACTAAGCTTAAACGACGCGTTAAACATTTC 50

## RESULT 7

AC008964/c  
LOCUS Homo sapiens chromosome 5 clone CTD-2360A17, complete sequence.  
DEFINITION AC008964 143841 bp DNA linear PRI 08-JUN-2001  
ACCESSION AC008964  
VERSION AC008964.6 GI:14329064  
KEYWORDS HTG.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 143841)  
AUTHORS DOE Joint Genome Institute and Stanford Human Genome Center.  
TITLE Direct Submission  
JOURNAL Unpublished

REFERENCE 2 (bases 1 to 143841)  
AUTHORS DOE Joint Genome Institute.  
TITLE Direct Submission  
JOURNAL Submitted (03-AUG-1999) Production Sequencing Facility, DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA

REFERENCE 3 (bases 1 to 143841)  
AUTHORS DOE Joint Genome Institute and Stanford Human Genome Center.  
TITLE Direct Submission  
JOURNAL Submitted (28-APR-2001) DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA

REFERENCE 4 (bases 1 to 143841)  
AUTHORS DOE Joint Genome Institute and Stanford Human Genome Center.  
TITLE Direct Submission  
JOURNAL Submitted (01-MAY-2001) DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA

REFERENCE 5 (bases 1 to 143841)  
AUTHORS DOE Joint Genome Institute and Stanford Human Genome Center.  
TITLE Direct Submission  
JOURNAL Submitted (08-JUN-2001) DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA

REFERENCE 6 (bases 1 to 143841)  
AUTHORS DOE Joint Genome Institute and Stanford Human Genome Center.  
TITLE Direct Submission  
JOURNAL Submitted (08-JUN-2001) DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA

COMMENT On Jun 8, 2001 this sequence version replaced gi:13876485.  
Draft Sequence Produced by DOE Joint Genome Institute  
www.jgi.doe.gov  
Finishing Completed at Stanford Human Genome Center  
www.shgc.stanford.edu  
Quality: Phrap Quality >=40 99.9% of Sequence;  
Estimated Total Number of Errors is 0.1.

STS Content:

SHGC-8813 GI4344.  
Location/Qualifiers  
1. 143841  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
/chromosome="5"  
/clone="CTD-2360A17"

## ORIGIN

Query Match 16.6%; Score 45.4; DB 8; Length 143841;  
Best Local Similarity 56.3%; Pred. No. 1.1;  
Matches 85; Conservative 0; Mismatches 66; Indels 0; Gaps 0;

QY 103 TGATGATGTCATTAGCATCAATCTCCACCGTCCCAATTTATTAGTTGTTGACAAATATCGA 162  
Db 18293 TAATGAAGTAGGAGCGCAAGGTTATCCATTGAAGTATTTATTTCACGCTCATCTTAA 18234

QY 163 CCGTCTAAGTTCCACACCGCGCTATAAGAGTTTCATTATAATTTTACGAAATAAAA 222  
Db 18233 GTGACAAATTCATACAGAAGACTATACAGAATCATATTTAATATATTTAAATTAAT 18174

QY 223 TCAGCAAAATTAATTTTCTTGACTAAGCTTA 253  
Db 18173 ACTTCAAATATCTTTTCACATTAAAGATGATTA 18143

## RESULT 8

## RESULT 9

AC010633/c  
LOCUS Homo sapiens chromosome 5 clone CTD-2317K6, complete sequence.  
DEFINITION AC010633 195931 bp DNA linear PRI 31-JUL-2001  
ACCESSION AC010633  
VERSION AC010633.8 GI:15042798  
KEYWORDS HTG.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 195931)  
AUTHORS DOE Joint Genome Institute and Stanford Human Genome Center.  
TITLE Direct Submission  
JOURNAL Unpublished

REFERENCE 2 (bases 1 to 195931)  
AUTHORS DOE Joint Genome Institute.  
TITLE Direct Submission  
JOURNAL Submitted (16-SEP-1999) Production Sequencing Facility, DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA

REFERENCE 3 (bases 1 to 195931)  
AUTHORS DOE Joint Genome Institute and Stanford Human Genome Center.  
TITLE Direct Submission  
JOURNAL Submitted (31-JUL-2001) DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA

REFERENCE 4 (bases 1 to 195931)  
AUTHORS DOE Joint Genome Institute and Stanford Human Genome Center.  
TITLE Direct Submission  
JOURNAL Submitted (31-JUL-2001) DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA

COMMENT On Jul 31, 2001 this sequence version replaced gi:13677000.  
Draft Sequence Produced by DOE Joint Genome Institute  
www.jgi.doe.gov  
Finishing Completed at Stanford Human Genome Center  
www.shgc.stanford.edu  
Quality: Phrap Quality >=40 99.6% of Sequence;  
Estimated Total Number of Errors is 0.4.

FEATURES  
source  
1. 195931  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
/chromosome="5"  
/clone="CTD-2317K6"

## ORIGIN

Query Match 16.6%; Score 45.4; DB 8; Length 195931;  
Best Local Similarity 56.3%; Pred. No. 1;  
Matches 85; Conservative 0; Mismatches 66; Indels 0; Gaps 0;

QY 103 TGATGATGTCATTAGCATCAATCTCCACCGTCCCAATTTATTAGTTGTTGACAAATATCGA 162  
Db 105243 TAATGAAGTAGGAGCGCAAGGTTATCCATTGAAGTATTTATTTCACGCTCATCTTAA 105184

QY 163 CCGTCTAAGTTCCACACCGCGCTATAAGAGTTTCATTATAATTTTACGAAATAAAA 222  
Db 105183 GTGACAAATTCATACAGAAGACTATACAGAATCATATTTAATATATTTAAATTAAT 105124

QY 223 TCAGCAAAATTAATTTTCTTGACTAAGCTTA 253  
Db 105123 ACTTCAAATATCTTTTCACATTAAAGATGATTA 105093

RESULT 9  
AC073652 167034 bp DNA linear HTG 26-JUN-2001  
LOCUS Homo sapiens chromosome 12 clone RP11-148D15, WORKING DRAFT  
DEFINITION AC073652  
ACCESSION AC073652  
VERSION AC073652.15 GI:14547458  
KEYWORDS HTG; HTGS PHASE1; HTGS DRAFT.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 167034)  
AUTHORS Muzny, D.M., Adams, C., Ali-Osman, F.R., Allen, C.,

REFERENCE 2 (bases 1 to 167034)  
AUTHORS Muzny, D.M., Adams, C., Ali-Osman, F.R., Allen, C.,

REFERENCE 3 (bases 1 to 167034)  
AUTHORS Muzny, D.M., Adams, C., Ali-Osman, F.R., Allen, C.,

REFERENCE 4 (bases 1 to 167034)  
AUTHORS Muzny, D.M., Adams, C., Ali-Osman, F.R., Allen, C.,

REFERENCE 5 (bases 1 to 167034)  
AUTHORS Muzny, D.M., Adams, C., Ali-Osman, F.R., Allen, C.,

Albrooks, S.L., Amaratunge, H.C., Are, J.R., Banks, T., Barbara, J., Benton, J., Bimaga, K., Blankenburg, K., Bonnin, D., Bouck, J., Bowie, S., Brieva, M., Brown, E., Bryant, N.P., Buhay, C., Burch, P., Burkett, C., Burrell, K.L., Byrd, N.C., Carron, T.F., Carter, M., Cavazos, S.R., Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Z., Chowdhry, I., Christopoulos, C., Cleveland, C.D., Cox, C., Coyle, M.D., Dathorne, S.R., David, R., Davila, M.L., Davis, C., Davy-Carroll, L., Dederich, D.A., Delaney, K.R., Delgado, O., Denn, A.L., Ding, Y., Dinh, H.H., Douthwaite, K.J., Draper, H., Dugan-Rocha, S., Durbin, K.J., Earnhart, C., Edgar, D., Edwards, C.C., Elhaj, C., Escotto, M., Falls, T., Ferraguto, D., Flagg, N., Ford, J., Foster, P., Frantz, P., Gabisi, A., Gao, J., Garcia, A., Garner, T., Garza, N., Gill, R., Gorrell, J.H., Guevara, W., Gunaratne, P., Hale, S., Hamilton, K., Harris, C., Harris, K., Hart, M., Havlak, P., Hawes, A., Hernandez, J., Hernandez, O., Hodgson, A., Hogue, M., Holloway, C., Hollins, B., Homs, F., Howard, S., Huber, J., Hulyk, S., Hume, J., Jackson, L.E., Jacobson, B., Jia, Y., Johnson, R., Jolivet, S., Joudah, S., Karlsson, E., Kelly, S., Khan, U., King, L., Korvah, J., Kovar, C., Kratoch, J., Kureshi, A., Landry, N., Leal, B., Lee, E., Lewis, L.C., Lewis, L., Li, J., Li, Z., Lichtarge, O., Lie, C., Liu, J., Liu, W., Lucier, R., Luna, R., Ma, J., Maheshwari, M., Mapua, P., Martin, R., Martindale, A., Martinez, E., Massey, E., Mawhiney, E., McLeod, M.P., Meador, M., Mei, G., Metzker, M., Miner, G., Miner, Z., Mitchell, T., Mohabbat, K., Morgan, M., Morris, S., Moser, M., Neal, D., Newton, J., Newton, N., Nguyen, A., Nguyen, N., Nguyen, N., Nickerson, E., Nwokenkwo, S., Oguh, M., Okwuonu, G., Oragunye, N., Oviedo, R., Pace, A., Payton, B., Peery, J., Perez, L., Peters, L., Pickens, R., Primus, E., Pu, L.L., Quiles, M., Ren, Y., Rives, M., Rojas, A., Rojibokan, I., Rolfe, M., Ruiz, S., Savery, G., Scherer, S., Scott, G., Shen, H., Shooshtari, N., Sison, I., Sodergren, E., Snaik, T., Sparks, A., Stanley, H., Stone, H., Sutton, A., Svatek, A., Tabor, P., Tamerisa, A., Tamerisa, K., Tang, H., Tansey, J., Taylor, C., Taylor, T., Telford, B., Thomas, N., Thomas, S., Umani, K., Vasquez, L., Vera, V., Villalón, D., Vinson, R., Wall, R., Wang, S., Ward-Moore, S., Warren, R., Washington, C., Watlington, S., Williams, G., Williamson, A., Wleczyk, R., Woodson, D., Worley, K., Wu, C., Wu, Y., Wu, Y.F., Zhou, J., Zorrilla, S., Nelson, D., Weinstein, G., and Gibbs, R.

Direct Submission  
Unpublished  
2 (bases 1 to 167034)  
Worley, K.C.

Direct Submission  
Submitted (28-JUN-2000) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA  
On Jun 25, 2001 this sequence version replaced gi:14150241.

----- Genome Center  
Center: Baylor College of Medicine  
Center code: BCM  
Web site: <http://www.hgsc.bcm.tmc.edu/>  
Contact: [hgsc-help@bcm.tmc.edu](mailto:hgsc-help@bcm.tmc.edu)  
----- Project Information  
Center project name: HBPB  
Center clone name: RP11-148D15  
----- Summary Statistics  
Sequencing vector: M13; L08821  
Chemistry: Dye-terminator Big Dye; 66% of reads  
Assembly program: Phrap; version 0.990329  
Consensus quality: 168964 bases at least Q40  
Consensus quality: 170984 bases at least Q30  
Estimated insert size: 167033; sum-of-contigs estimation  
Quality coverage: 0x in Q20 bases; agarose-gel estimation  
Quality coverage: 6x in Q20 bases; sum-of-contigs estimation

----- NOTE: Estimated insert size may differ from sequence length  
(see [http://www.hgsc.bcm.tmc.edu/docs/genbank\\_draft\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html)).  
\* NOTE: This is a 'working draft' sequence. It currently consists of 1 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as

\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.  
\* 1 167034: contig of 167034 bp in length.

FEATURES  
Location/Qualifiers  
1..167034  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
/chromosome="12"  
/clone="RP11-148D15"

ORIGIN

Query Match 16.1%; Score 44.2; DB 14; Length 167034;  
Best Local Similarity 51.2%; Pred. No. 2;  
Matches 103; Conservative 0; Mismatches 98; Indels 0; Gaps 0;

QY 48 ATTCCAAAAGTCAATAAAGTGTGACGTGATGATCTTACGCTTTAAACATCGCATGATG 107  
DB 92792 ATTATCAAAATATACATAGAAATTTGATTTCTAAATAGCAATATACATTTTG 92851

QY 108 ATGTCATTAGCATCAATCTCCACCGTCCCAATTTATTAGTTGTTGCAATATCGACCGTC 167  
DB 92852 AATTCATTTTTTAAACCACTACACTATTTCAGTGTATATGTGTCAGAAACATTTATAAG 92911

QY 168 TTAGTTCCACCGGAGGCTATAGAGTTTCATTATAATTTTACCAAAATATAATCAGC 227  
DB 92912 GATAGTTTCATCCATGAAAGATGAAAAATTTGATCATGAAGTACTACTATAAAAGT 92971

QY 228 AATTAATTTTTTCTTGACTAA 248  
DB 92972 CATTTCTTTTATTGTTTAA 92992

RESULT 10  
AC073528/c

LOCUS AC073528 167041 bp DNA linear PRI 01-MAY-2001  
DEFINITION Homo sapiens 12 BAC RP11-8P16 (Roswell Park Cancer Institute Human BAC Library) complete sequence.  
ACCESSION AC073528  
VERSION AC073528.26 GI:13899199  
KEYWORDS HTG.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Homnidae; Homo.

REFERENCE  
1 (bases 1 to 167041)  
Muzny, D.M., Adams, C., Adio-Oduola, B., Ali-ozman, F.R., Allen, C., Albrooks, S.L., Amaratunge, H.C., Are, J.R., Banks, T., Barbari, J., Benton, J., Bimaga, K., Blankenburg, K., Bonnin, D., Bouck, J., Bowie, S., Brieva, M., Brown, E., Bryant, N.P., Buhay, C., Burch, P., Burkett, C., Burrell, K.L., Byrd, N.C., Carron, T.F., Carter, M., Cavazos, S.R., Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Z., Chiu, D., Chowdhry, I., Christopoulos, C., Cleveland, C.D., Cox, C., Coyle, M.D., Dathorne, S.R., David, R., Davila, M.L., Davis, C., Davy-Carroll, L., Dederich, D.A., Delaney, K.R., Delgado, O., Denn, A.L., Ding, Y., Dinh, H.H., Douthwaite, K.J., Draper, H., Dugan-Rocha, S., Durbin, K.J., Earnhart, C., Edgar, D., Edwards, C.C., Elhaj, C., Emerling, S., Escotto, M., Falls, T., Ferraguto, D., Flagg, N., Ford, J., Foster, P., Frantz, P., Gabisi, A., Gao, J., Garcia, A., Garner, T., Garza, N., Gill, R., Gorrell, J.H., Guevara, W., Gunaratne, P., Hale, S., Hamilton, K., Han, J., Harris, C., Harris, K., Hart, M., Havlak, P., Hawes, A., Hernandez, J., Hernandez, O., Hodgson, A., Hogue, M., Holloway, C., Hollins, B., Homs, F., Howard, S., Huber, J., Hulyk, S., Hume, J., Johnson, R., Jolivet, S., Joudah, S., Jacobson, B., Jia, Y., Johnson, R., Jolivet, S., Joudah, S., Karlsson, E., Kelly, S., Khan, U., King, L., Korvah, J., Kovar, C., Kratoch, J., Kureshi, A., Landry, N., Leal, B., Lee, E., Lewis, L.C., Lewis, L., Li, J., Li, Z., Lichtarge, O., Lie, C., Liu, J., Liu, W., Lucier, R., Luna, R., Ma, J., Maheshwari, M., Mapua, P., Martin, R., Martindale, A., Martinez, E., Massey, E., Mawhiney, E., McLeod, M.P., Meador, M., Mei, G., Metzker, M., Miner, G., Miner, Z., Mitchell, T., Mohabbat, K., Morgan, M., Morris, S., Moser, M., Neal, D., Newton, J., Newton, N., Nguyen, A., Nguyen, N., Nguyen, N., Nickerson, E., Nwokenkwo, S., Oguh, M., Okwuonu, G., Oragunye, N., Oviedo, R., Pace, A., Payton, B., Peery, J., Perez, L., Peters, L., Pickens, R., Primus, E., Pu, L.L., Quiles, M., Ren, Y., Rives, M., Rojas, A., Rojibokan, I., Rolfe, M., Ruiz, S., Savery, G., Scherer, S., Scott, G., Shen, H., Shooshtari, N., Sison, I., Sodergren, E., Snaik, T., Sparks, A., Stanley, H., Stone, H., Sutton, A., Svatek, A., Tabor, P., Tamerisa, A., Tamerisa, K., Tang, H., Tansey, J., Taylor, C., Taylor, T., Telford, B., Thomas, N., Thomas, S., Umani, K., Vasquez, L., Vera, V., Villalón, D., Vinson, R., Wall, R., Wang, S., Ward-Moore, S., Warren, R., Washington, C., Watlington, S., Williams, G., Williamson, A., Wleczyk, R., Woodson, D., Worley, K., Wu, C., Wu, Y., Wu, Y.F., Zhou, J., Zorrilla, S., Nelson, D., Weinstein, G., and Gibbs, R.





```

/rpt_family="AluYa5"
7924..7941
/rpt_family="AT_rich"
complement(8256..8962)
/rpt_family="L2"
8966..9036
/rpt_family="MIR"
complement(9803..10024)
/rpt_family="L2"
11288..11321
/rpt_family="AT_rich"
11478..12092
/rpt_family="L2"
complement(12182..12303)
/rpt_family="MLT1G"
complement(12398..12636)
/rpt_family="MLT1G"
complement(12641..13107)

Query Match      16.1%; Score 44.2; DB 8; Length 167041;
Best Local Similarity 51.2%; Pred. No. 2;
Matches 103; Conservative 0; Mismatches 98; Indels 0; Gaps 0;

QY 48 ATTCCAAAGTCAATAAGTGTGACGTGATGATCTAGCGTTTAAACATCGCATGATG 107
Db 74249 ATTACAAATATACATAGAAATTTGATCTTAAATAGCAATATAACATCACATTTTG 74190

QY 108 ATGTCATTAGCATCAATCTCCACCGTCCCAATTTATTAGTTGTCACATATCGACCGTC 167
Db 74189 AATTGATTTTAAACCACTACACTATTCTAGTGTATATGTGCGAGGAACATTTATAAG 74130

QY 168 TAAGTTCACACGACGGCTTAAGAGTTTCATTATAAATTTTAGCAAAATAAATCAGC 227
Db 74129 GATAGTTCAATCCGATGAAGATGAAATTTGATCATGAAGATTACTACTATAAAAGT 74070

QY 228 AATAATTTTCTTGACTAA 248
Db 74069 CATTCTTTTATTGTTTAA 74049

RESULT 11
AC024628 207976 bp DNA linear HTG 25-MAY-2000
LOCUS Homo sapiens chromosome 12 clone RP11-632011 map 12, WORKING DRAFT
DEFINITION SEQUENCE, 24 unordered pieces.
ACCESSION AC024628
VERSION HTG; HTGS_PHASE1; HTGS_DRAFT.
KEYWORDS Homo sapiens (human)
SOURCE Homo sapiens
ORGANISM

REFERENCE
1 (bases 1 to 207976)
AUTHORS Birren,B., Linton,L., Nusbaum,C. and Lander,E.
TITLE Homo sapiens chromosome 12, clone RP11-632011
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 207976)
AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
Anderson,S., Baldwin,J., Barna,N., Baetien,V., Beda,F.,
Boguslavsky,L., Bouckgaeter,B., Brown,A., Burkett,G.,
Campopiano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S.,
Collamore,A., Cooke,P., DeArrellano,K., Dewar,K., Diaz,J.S.,
Dodge,S., Domino,M., Doyle,M., Ferreira,P., FitzHugh,W., Gage,D.,
Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,I., Karatas,A.,
Klein,J., LaRocque,K., Lamazares,R., Landers,T., Lehoczy,J.,
Levine,R., Liu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N.,
McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheeters,R.,
Melarim,J., Meneus,L., Mihova,T., Miranda,C., Mienga,V., Morrow,J.,
Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
O'Neill,D., Olivarez,M., Oliver,J., Peterson,K., Pierre,N.,

```

```

Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,
Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,
Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
Tefaye,S., Theodore,J., Tirrell,A., Travers,M., Trigglio,J.,
Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
Young,G., Zainoun,J., Zimmer,A. and Zody,M.
Direct Submission
Submitted (01-MAR-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On May 25, 2000 this sequence version replaced gi:7144995.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L5991
Center clone name: 632_O_11
----- Summary Statistics
Sequencing vector: M13; M77815; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 193853 bases at least Q40
Consensus quality: 200610 bases at least Q30
Consensus quality: 203401 bases at least Q20
Insert size: 228000; agarose-fp
Quality coverage: 4.1 in Q20 bases; agarose-fp
Quality coverage: 4.5 in Q20 bases; sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 24 contigs. The true order of the contigs
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence.
* as soon as it is available and the accession number will
* be preserved.
* 1 1209: contig of 1209 bp in length
* 1210 1309: gap of 100 bp
* 1310 2947 1209: contig of 1637 bp in length
* 2947 3047 5028: gap of 100 bp
* 3047 5029 5128: contig of 1982 bp in length
* 5029 5129 7145: gap of 100 bp
* 5129 7146 7245: contig of 2017 bp in length
* 7146 7246 8721: gap of 100 bp
* 7246 8722 8821: contig of 1476 bp in length
* 8722 8822 11031: gap of 100 bp
* 8822 11031 11131: contig of 2210 bp in length
* 11031 11132 12855: gap of 100 bp
* 11132 12855 12955: contig of 1724 bp in length
* 12855 12956 15225: gap of 100 bp
* 12956 15225 16025: contig of 2970 bp in length
* 15225 16026 18959: gap of 100 bp
* 16026 18959 19059: contig of 2934 bp in length
* 18959 19060 22533: gap of 100 bp
* 19060 22533 24536: contig of 3474 bp in length
* 22533 24536 27457: gap of 100 bp
* 24536 27457 29768: contig of 4823 bp in length
* 27457 29768 29869: gap of 100 bp
* 29768 29869 34812: contig of 2212 bp in length
* 29869 34812 34913: gap of 100 bp
* 34812 34913 41011: contig of 4944 bp in length
* 34913 41011 45237: gap of 100 bp
* 41011 45237 45337: contig of 6039 bp in length
* 45237 45337 55231: gap of 100 bp
* 45337 55231 55330: contig of 4125 bp in length
* 55231 55330: gap of 100 bp
* 55330: contig of 9894 bp in length
* 55330: gap of 100 bp

```

TITLE  
JOURNAL  
COMMENT



```

* 55331 65620: contig of 10290 bp in length
* 65621 65720: gap of 100 bp
* 76314: contig of 10594 bp in length
* 76315 76414: gap of 100 bp
* 76415 88196: contig of 11782 bp in length
* 88197 88296: gap of 100 bp
* 88297 102065: contig of 13769 bp in length
* 102066 102165: gap of 100 bp
* 102166 116370: contig of 14205 bp in length
* 116371 116470: gap of 100 bp
* 116471 133517: contig of 17047 bp in length
* 133518 133617: gap of 100 bp
* 133618 171789: contig of 38172 bp in length
* 171790 171899: gap of 100 bp
* 171899 207976: contig of 36087 bp in length.

```

# FEATURES

```

source
1. .207976
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="12"
/map="12"
/clone="RP11-632011"
/clone_lib="RPC1-11 Human Male BAC"

misc_feature
1. .1209
/notes="assembly_fragment"
gap
1210. .1309
/estimated_length=100
misc_feature
1310. .2946
/notes="assembly_fragment"
gap
2947. .3046
/estimated_length=100
misc_feature
3047. .5028
/notes="assembly_fragment"
gap
5029. .5128
/estimated_length=100
misc_feature
5129. .7145
/notes="assembly_fragment"
gap
7146. .7245
/estimated_length=100
misc_feature
7246. .8721
/notes="assembly_fragment"
gap
8722. .8821
/estimated_length=100
misc_feature
8822. .11031
/notes="assembly_fragment"
gap
11032. .11131
/estimated_length=100
misc_feature
11132. .12855
/notes="assembly_fragment"
gap
12856. .12955
/estimated_length=100
misc_feature
12956. .15925
/notes="assembly_fragment"
gap
15926. .16025
/estimated_length=100
misc_feature
16026. .18959
/notes="assembly_fragment"
gap
18960. .19059
/estimated_length=100
misc_feature
19060. .22533
/notes="assembly_fragment"
gap
22534. .22633
/estimated_length=100
misc_feature
22634. .27456
/notes="assembly_fragment"
gap
27457. .27556
/estimated_length=100
misc_feature
27557. .29768
/notes="assembly_fragment"
clone_end:SP6
vector_side:right
29769. .29868

```

```

misc_feature
29869. .34812
/notes="assembly_fragment"
gap
34813. .34912
/estimated_length=100
misc_feature
34913. .41011
/notes="assembly_fragment"
gap
41012. .41111
/estimated_length=100
misc_feature
41112. .45236
/notes="assembly_fragment"
gap
45237. .45336
/estimated_length=100
misc_feature
45337. .55230
/notes="assembly_fragment"
gap
55231. .55330
/estimated_length=100
misc_feature
55331. .65620
/notes="assembly_fragment"
gap
65621. .65720
/estimated_length=100
misc_feature
65721. .76314
/notes="assembly_fragment"

```

```

Query Match 16.1%; Score 44.2; DB 14; Length 207976;
Best Local Similarity 51.2%; Pred.No.1.9;
Matches 103; Conservative 0; Mismatches 98; Indels 0; Gaps 0;

QY 48 ATTCCAAAAGTCAATAAAGTGCATGATACATCTAGCTTTAAACATCCATGATG 107
DB 145560 ATTATACAAATATACACATAAGAAATTTGATCTTAATAAGCAATATAACATCATTTTG 145619

QY 108 ATGTCAATAGCATCAATCTCCACGTCCTCAATTTATTAGTTGTGACAATATCGACGTC 167
DB 145620 AATTGATTTTAAACCAACTACATCTATTCAGTGTATATGTCAGGAGAACTTTATAAAG 145679

QY 168 TAAGTTCCACACGCGCTATAGAGTTTCATATATAAATTTTACGAAATAAAATCAGC 227
DB 145680 GATAGTTTCATCCCATGAAAGATGAAAATTTGATCATGAGAAATTTACTAATAAAAGT 145739

QY 228 AATAATTTTTCCTTCACTAA 248
DB 145740 CATTCTTTTATTGTTTAA 145760

```

# RESULT 12

```

AC005158 Homo sapiens BAC clone GS1-250N6 from 7, complete sequence.
LOCUS AC005158 266344 bp DNA linear PRI 30-JAN-2004
DEFINITION Homo sapiens BAC clone GS1-250N6 from 7, complete sequence.
ACCESSION AC005158
VERSION AC005158.3 GI:13446342
KEYWORDS HTG.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominiidae; Homo.
REFERENCE
1 (bases 1 to 266344)
AUTHORS Hillier,L.W., Fulton,R.S., Fulton,L.A., Graves,T.A., Pepin,K.H.,
Wagner-McPherson,C., Layman,D., Maas,J., Jaeger,S., Walker,R.,
Wylie,K., Sekhon,M., Becker,M.C., O'Laughlin,M.D., Schaller,M.E.,
Pewell,G.A., Delehaanty,K.D., Miner,T.L., Nash,W.E., Cordes,M.,
Du,H., Sun,H., Edwards,J., Bradshaw-Cordum,H., Ali,J., Andrews,S.,
Leak,A., Vanbrunt,A., Nguyen,C., Du,F., Lamaz,B., Courtney,L.,
Kalicki,J., Ozersky,P., Bielicki,L., Scott,K., Holmes,A.,
Harkins,R., Harris,A., Strong,C.M., Hou,S., Tomlinson,C.,
Dauphin-Kohlberg,S., Kozlowski-Reilly,A., Leonard,S., Rohlfing,T.,
Rock,S.M., Tin-Wollam,A.M., Abbott,A., Minx,P., Maupin,R.,
Strommatt,C., Latreille,P., Miller,N., Johnson,D., Murray,J.,
Woessner,J.P., Wendt,M.C., Yang,S.P., Schultz,B.R., Wallis,J.W.,
Spieth,J., Bieri,T.A., Nelson,J.O., Berkowicz,N., Wohlmann,P.E.,
Cook,L.L., Hickenbotham,M.T., Eldred,J., Williams,D., Bedell,J.A.,
Mardis,E.R., Clifton,S.W., Chissole,S.L., Marra,M.A., Raymond,C.,

```

Haugen, E., Gillett, W., Zhou, Y., James, R., Phelps, K., Iadanoto, S., Bub, K., Simms, E., Levy, R., Clendenning, J., Kaul, R., Kent, W.J., Furey, T.S., Baertsch, R.A., Brent, M.R., Keibler, E., Flicek, P., Bork, P., Suyama, M., Bailey, J.A., Portnoy, M.E., Torrents, D., Chinwalla, A.T., Gish, W.R., Eddy, S.R., McPherson, J.D., Olson, M.V., Eichler, E.E., Green, E.D., Waterston, R.H. and Wilson, R.K.  
The DNA sequence of human chromosome 7  
Nature 424 (6945), 157-164 (2003)  
12853948  
2 (bases 1 to 266344)  
Bauer, C., McPherson, C. and Williams, D.  
The sequence of Homo sapiens BAC clone GS1-250N6  
Unpublished (2001)  
3 (bases 1 to 266344)  
Waterston, R.H.  
Direct Submission  
Submitted (20-JUN-1998) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA  
4 (bases 1 to 266344)  
Waterston, R.H.  
Direct Submission  
Submitted (17-JUN-1999) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA  
5 (bases 1 to 266344)  
Waterston, R.  
Direct Submission  
Submitted (30-SEP-2000) Department of Genetics, Washington University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA  
6 (bases 1 to 266344)  
Waterston, R.H.  
Direct Submission  
Submitted (25-MAR-2001) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA  
7 (bases 1 to 266344)  
Waterston, R.  
Direct Submission  
Submitted (04-JUN-2002) Department of Genetics, Washington University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA  
8 (bases 1 to 266344)  
Wilson, R.  
Direct Submission  
Submitted (30-JAN-2004) Department of Genetics, Washington University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA  
On Mar 25, 2001 this sequence version replaced gi:5091650.  
----- Genome Center  
Center: Washington University Genome Sequencing Center  
Center code: WUGSC  
Web site: <http://genome.wustl.edu>  
Contact: [sapiens@wustl.edu](mailto:sapiens@wustl.edu)  
----- Summary Statistics  
Center project name: H\_GS250N06  
-----

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:  
The sequence of this clone was established as part of a mapping and sequencing collaboration between the NHGRI Chromosome 7 Mapping Project (Eric D. Green, Director), John D. McPherson in the

Department of Genetics (Washington University), and the Washington University Genome Sequencing Center. For additional information about the map position of this sequence, see <http://www.nhgri.nih.gov/DIR/OTB/CHR7>, send [mailto:egreen@nhgri.nih.gov](mailto:mailto:egreen@nhgri.nih.gov), or see <http://genome.wustl.edu>

SOURCE INFORMATION:  
This clone is from the first BAC library from Genome Systems, Inc. (<http://www.genomesystems.com>).  
Cell line: lymphoblastoid

Haplotypes: two  
VECTOR: pBelcBAC  
Selection: chloramphenicol

NEIGHBORING SEQUENCE INFORMATION:  
Actual start of this clone is at base position 1 of GS1-250N6 actual end is at base position 266344 of GS1-250N6.

FEATURES	Source
repeat_region	1..266344 /organism="Homo sapiens" /mol_type="genomic DNA" /db_xref="taxon:9606" /chromosome="7" /map="7" /clone="GS1-250N6" /clone_lib="GSBAC1" 762..902 /rpt_family="MIR"
repeat_region	3681..3865 /rpt_family="MER1_type"
repeat_region	4657..4713 /rpt_family="(TA)n"
repeat_region	5001..5125 /rpt_family="MIR"
repeat_region	6297..6466 /rpt_family="MER1_type"
repeat_region	10202..10520 /rpt_family="Alu"
repeat_region	11275..11427 /rpt_family="MIR"
repeat_region	11592..11824 /rpt_family="MIR"
repeat_region	12004..12103 /rpt_family="Alu"
repeat_region	12899..13675 /rpt_family="L1"
repeat_region	14550..14874 /rpt_family="Alu"
repeat_region	16275..16580 /rpt_family="Alu"
repeat_region	18336..18380 /rpt_family="CT-rich"
repeat_region	19426..19681 /rpt_family="MIR"
repeat_region	19691..19827 /rpt_family="L2"
repeat_region	19857..20139 /rpt_family="L2"
repeat_region	20140..20482 /rpt_family="MER2_type"
repeat_region	20483..20629 /rpt_family="L2"
repeat_region	21873..22161 /rpt_family="Alu"
repeat_region	22458..22842 /rpt_family="L2"
repeat_region	23395..23447 /rpt_family="GA-rich"
repeat_region	23875..24166 /rpt_family="Alu"
repeat_region	25208..25431 /rpt_family="Alu"
repeat_region	28192..28253

```
repeat_region /rpt_family="GA-rich"
28282..28508
repeat_region /rpt_family="MIR"
30611..30652
gene /rpt_family="(CA)n"
complement(30701..105771)
/gene="GLI3"
complement(join(30701..30943,105605..105771))
/gene="GLI3"
complement(join(<30701..30943,105605..1057728))
/gene="GLI3"
/note="Homo sapiens GLI-Kruppel family member GLI3 (Greig
cephalopolysyndactyly syndrome) (GLI3), mRNA.;
H_GS250N06.1
This gene was based on gi(13518031)
Continues as H_GS489L14.1"
/codon_start=1
/product="unknown"
/protein_id="AA02015.1"
/db_xref="GI:41393490"
/translation="HEAQSHTTTEKKKVENSVKCSRTDVSEKAVASSTTSNDE
SPGQTHRRRAITMPQNVGLSKVSEPTSSDERASLIKKEIHGSLPHVAPSV
PYRGTVFAMDPNGYMEPHY"
repeat_region 34549..34579
/rpt_family="(TTTTG)n"
repeat_region 34931..34953
/rpt_family="(TCTCCC)n"
repeat_region 36317..36368

Query Match 15.8%; Score 43.4; DB 8; Length 266344;
Best Local Similarity 50.2%; Pred. No. 2.7;
Matches 107; Conservative 0; Mismatches 106; Indels 0; Gaps 0;

QY 22 AATTTTCCACATGCGATCTCTTTATTTCCAAAAGTCATAAAGTGTGACGTGATGATA 81
Db 162852 AAAATATCTGAAATATCTTCTTTATTGAGAATTTCAAAAATTAAGATCTCAAAAAC 162911
QY 82 CTTACGCTTTAAACATCCGATCATGTCATTAGCATCAATCTCCACCGTCCCAATTTA 141
Db 162912 AGCATAGTTAGTAGAAGAATAAATATCATTTACAGATGAATTTCTAATGTCAAAATA 162971

QY 142 TTTAGTTGTTGACAATATCAGCGTCTTAAGTTCCACACCGCGCTATAGAGTTTTCATT 201
Db 162972 ATTCTTAAAGATTTCCACAAATGAAACACATAATATCCGCTTGAAACCTCAIT 163031
QY 202 ATAAATTTTACAAAATAAATAACAGCAATAAT 234
Db 163032 TTATATTTGCCAAAATAAGAACAGAGGAAT 163064

RESULT 13
AC008563
LOCUS
DEFINITION Homo sapiens chromosome 5 clone CTC-536A23, linear HTG 18-JUL-2000
7 ordered pieces.
AC008563
VERSION AC008563.4 GI:7711299
KEYWORDS HTG; HTGS_PHASE2; HTGS_DRAFT.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryote; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Homnidae; Homo.
REFERENCE
1 (bases 1 to 94703)
DOE Joint Genome Institute.
TITLE Sequencing of Human Chromosome 5
JOURNAL Unpublished
2 (bases 1 to 94703)
DOE Joint Genome Institute.
AUTHORS DOE Joint Genome Institute.
TITLE Direct Submission
JOURNAL Submitted (03-AUG-1999) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
COMMENT On May 6, 2000 this sequence version replaced gi.7708951.
```

```
-----Genome Center
Center: Joint Genome Institute
Center Code: JGI
Web site: http://www.jgi.doe.gov
-----
Project Information
Center Project Name: 391027, H433
Center clone name: CIT-HSPC_536A23
-----
Summary Statistics
Consensus quality: 88918 bases at least Q40
Consensus quality: 93000 bases at least Q30
Consensus quality: 93760 bases at least Q20
Estimated insert size: 99000; pulse field gel estimation
Estimated insert size: 94453; sum-of-contigs estimation
Quality coverage: 5.84 in Q20 bases; pulse field gel estimation.
Quality coverage: 6.12 in Q20 bases; sum-of-contigs estimation.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 7 contigs. Gaps between the contigs
* are represented as runs of N. The order of the pieces
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have
* provided by the submitter.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
* 1 34594: contig of 34594 bp in length
* 34595 34694: gap of unknown length
* 34695 51143: contig of 16449 bp in length
* 51144 51243: gap of unknown length
* 51244 66425: contig of 15182 bp in length
* 66426 66525: gap of unknown length
* 66526 72366: contig of 5841 bp in length
* 72367 72467: gap of unknown length
* 72467 79229: contig of 7163 bp in length
* 79229 79729: gap of unknown length
* 79730 80679: contig of 950 bp in length
* 80680 80779: gap of unknown length
* 80780 94703: contig of 13924 bp in length.

FEATURES
source
1..94703
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="5"
/clone="CTC-536A23"
/clone_lib="Caltech human BAC library C"
34595..34694
/estimated_length=unknown
51144..51243
/estimated_length=unknown
66426..66525
/estimated_length=unknown
72367..72466
/estimated_length=unknown
79630..79729
/estimated_length=unknown
80680..80779
/estimated_length=unknown

ORIGIN
Query Match 15.5%; Score 42.4; DB 14; Length 94703;
Best Local Similarity 48.4%; Pred. No. 5.8;
Matches 118; Conservative 0; Mismatches 126; Indels 0; Gaps 0;

QY 4 AGAATATATAGCCGATACCAATTTTCCACATGACCTTCCTTATTCAAAAGTCAATA 63
Db 72000 AGTCTTAATACAAAGATTAATAATTTTAGATGGTGATATAATTTTCAAAAAGTAACC 72059
QY 64 AAGTGTGACGTGATGATCTAGCTTTTAAACATCGCATGATGATGATGATGATGATGAT 123
Db 72060 AATGCTGGGGTTATTAGGAATTCGATTTAGTTACTGACAGGATGATATTATGATG 72119
```



GenCore version 5.1.6  
Copyright (c) 1993 - 2006 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 12, 2006, 00:23:59 ; Search time 812 Seconds  
(without alignments)  
2790.406 Million cell updates/sec

Title: US-10-800-161-28

Perfect score: 274  
Sequence: 1 tctagaataatagccgatac.....acgacgccgtaacattttc 274

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 9793542 seqs, 4134689005 residues

Total number of hits satisfying chosen parameters: 19587084

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications NA.Main.\*

- 1: /cgn2\_6/ptodata/1/pubpna/US07\_PUBCOMB.seq.\*
- 2: /cgn2\_6/ptodata/1/pubpna/US08\_PUBCOMB.seq.\*
- 3: /cgn2\_6/ptodata/1/pubpna/US09A\_PUBCOMB.seq.\*
- 4: /cgn2\_6/ptodata/1/pubpna/US09B\_PUBCOMB.seq.\*
- 5: /cgn2\_6/ptodata/1/pubpna/US10A\_PUBCOMB.seq.\*
- 6: /cgn2\_6/ptodata/1/pubpna/US10B\_PUBCOMB.seq.\*
- 7: /cgn2\_6/ptodata/1/pubpna/US10C\_PUBCOMB.seq.\*
- 8: /cgn2\_6/ptodata/1/pubpna/US10D\_PUBCOMB.seq.\*
- 9: /cgn2\_6/ptodata/1/pubpna/US10E\_PUBCOMB.seq.\*
- 10: /cgn2\_6/ptodata/1/pubpna/US11\_PUBCOMB.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	274	100.0	274	7	US-10-800-161-28
2	272.4	99.4	274	7	Sequence 28, Appl
3	272.4	99.4	544	7	Sequence 26, Appl
4	272.4	99.4	862	7	Sequence 27, Appl
5	272.4	99.4	962	7	Sequence 25, Appl
6	272.4	99.4	1700	7	Sequence 24, Appl
7	272.4	99.4	1700	8	Sequence 3, Appl
8	52.8	19.3	509	7	Sequence 3, Appl
9	52.8	19.3	509	8	Sequence 1, Appl
10	43.4	15.8	2469	5	Sequence 102510,
11	43.4	15.8	2469	5	Sequence 102511,
12	43.4	15.8	2469	5	Sequence 112106,
13	43.4	15.8	2469	6	Sequence 102510,
14	43.4	15.8	2469	6	Sequence 102511,
15	43.4	15.8	2469	6	Sequence 112106,
16	43	15.7	321	6	Sequence 676, App
17	42	15.3	229	3	Sequence 1719, App
18	42	15.3	229	5	Sequence 1719, App
19	42	15.3	1197	6	Sequence 6, Appl
20	42	15.3	2026	6	Sequence 22, Appl
21	42	15.3	4673	6	Sequence 542, App
22	39.8	14.5	195	7	Sequence 36002, A
23	39.8	14.5	195	7	Sequence 36002, A

24	39.4	14.4	375	3	US-09-918-995-37451	Sequence 37451, A
25	38.8	14.2	6301	6	US-10-311-455-26	Sequence 26, Appl
26	38.4	14.0	794	5	US-10-027-632-147687	Sequence 147687,
27	38.4	14.0	794	5	US-10-027-632-147688	Sequence 147688,
28	38.4	14.0	794	5	US-10-027-632-147689	Sequence 147689,
29	38.4	14.0	794	6	US-10-027-632-147687	Sequence 147687,
30	38.4	14.0	794	6	US-10-027-632-147688	Sequence 147688,
31	38.4	14.0	794	6	US-10-027-632-147689	Sequence 147689,
32	37.8	13.8	319608	6	US-10-147-603-1	GENERAL INFORMATI
33	37.6	13.7	513	4	US-09-925-065A-534423	Sequence 534423,
34	37.2	13.6	462	7	US-10-242-535A-7701	Sequence 7701, Ap
35	37.2	13.6	462	7	US-10-085-783A-7701	Sequence 2, Appl
36	37.2	13.6	1852	3	US-09-813-356-2	Sequence 1, Appl
37	37.2	13.6	1963	3	US-09-813-356-1	Sequence 3, Appl
38	37.2	13.6	1963	3	US-09-813-356-3	Sequence 585231,
39	36	13.1	509	4	US-09-925-065A-585231	Sequence 47, Appl
40	36	13.1	182328	7	US-10-235-192A-47	Sequence 6, Appl
41	35.6	13.0	87394	8	US-10-810-788A-6	Sequence 69, Appl
42	35.4	12.9	140040	7	US-10-275-762-69	Sequence 2, Appl
43	35.4	12.9	367378	6	US-10-312-841-2	Sequence 10963, A
44	35.2	12.8	621	7	US-10-282-122A-10963	Sequence 2117, Ap
45	35	12.8	516	6	US-10-029-386-2117	

## ALIGNMENTS

### RESULT 1

US-10-800-161-28  
; Sequence 28, Application US/10800161  
; Publication No. US20040154051A1  
; GENERAL INFORMATION:  
; APPLICANT: Cade, Rebecca M  
; APPLICANT: Lawton, Kay Ann  
; APPLICANT: Dietrich, Robert A  
; TITLE OF INVENTION: INDUCIBLE PROMOTERS  
; FILE REFERENCE: A-31089CIP1  
; CURRENT APPLICATION NUMBER: US/10/800,161  
; CURRENT FILING DATE: 2004-03-12  
; PRIOR APPLICATION NUMBER: 60/171,008  
; PRIOR FILING DATE: 1999-12-15  
; PRIOR APPLICATION NUMBER: 60/175,519  
; PRIOR FILING DATE: 2000-01-11  
; NUMBER OF SEQ ID NOS: 31  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 28  
; LENGTH: 274  
; TYPE: DNA  
; ORGANISM: Arabidopsis thaliana  
US-10-800-161-28

Query Match	100.0%;	Score 274;	DB 7;	Length 274;
Best Local Similarity	100.0%;	Pred. No. 8.1e-61;	Mismatches 0;	Indels 0;
Matches 274;	Conservative 0;			
Qy	1	TCTAGAATAATAGCCGATACCAATTTTCCACACATGACATTCCTTTATTTCCAAAAGTCA	60	
Db	1	TCTAGAATAATAGCCGATACCAATTTTCCACACATGACATTCCTTTATTTCCAAAAGTCA	60	
Qy	61	ATAAAGTGTCACGTCATGATACCTTACGCTTTTAAACATCGCATGATGATGATGATGAT	120	
Db	61	ATAAAGTGTCACGTCATGATACCTTACGCTTTTAAACATCGCATGATGATGATGATGAT	120	
Qy	121	CAATCTCACCGTCCAAATTTATTTAGTTGTGACAAATCGACCGTCTAAGTTCCACACC	180	
Db	121	CAATCTCACCGTCCAAATTTATTTAGTTGTGACAAATCGACCGTCTAAGTTCCACACC	180	
Qy	181	GACCGCTATAGAGTTTCATTTATATAAATTTAGCAAAATAAATCAGCAAAATAATTTTTC	240	
Db	181	GACCGCTATAGAGTTTCATTTATATAAATTTAGCAAAATAAATCAGCAAAATAATTTTTC	240	
Qy	241	TTGACTAAGCTTAAACGACGCGTTTAAACATTTTC	274	
Db	241	TTGACTAAGCTTAAACGACGCGTTTAAACATTTTC	274	



```
|||||
Db 818 TTGACTAAGCTTAAACGACGCCGTTAAACATTTTC 851

RESULT 5
US-10-800-161-24
; Sequence 24, Application US/10800161
; Publication No. US20040154051A1
; GENERAL INFORMATION:
; APPLICANT: Cade, Rebecca M
; APPLICANT: Dietrich, Robert A
; APPLICANT: Lawton, Kay Ann
; TITLE OF INVENTION: INDUCIBLE PROMOTERS
; FILE REFERENCE: A-31089CIP1
; CURRENT APPLICATION NUMBER: US/10/800,161
; CURRENT FILING DATE: 2004-03-12
; PRIOR APPLICATION NUMBER: 60/171,008
; PRIOR FILING DATE: 1999-12-15
; PRIOR APPLICATION NUMBER: 60/175,519
; PRIOR FILING DATE: 2000-01-11
; NUMBER OF SEQ ID NOS: 31
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 24
; LENGTH: 962
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
US-10-800-161-24

Query Match 99.4%; Score 272.4; DB 7; Length 962;
Best Local Similarity 99.6%; Pred. No. 3.3e-60;
Matches 273; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 TCTAGAAATATAGCGGATACCAATTTTCCACATGAGCTTCCTTTATTCCTCAAAAGTCA 60
Db 678 TCTAAAAATATAGCGGATACCAATTTTCCACATGAGCTTCCTTTATTCCTCAAAAGTCA 737

Qy 61 ATAAAGTGTGACGTCATGATACCTTACGCTTTAAACATCGCATGATGTCATTAGCAT 120
Db 738 ATAAAGTGTGACGTCATGATACCTTACGCTTTAAACATCGCATGATGTCATTAGCAT 797

Qy 121 CAATCTCCACCGTCCAAATTTATTTAGTTGTTGACAATATCGACCGTCTTAAGTTCACAC 180
Db 798 CAATCTCCACCGTCCAAATTTATTTAGTTGTTGACAATATCGACCGTCTTAAGTTCACAC 857

Qy 181 GACGGCTATAAGAGTTTCATTTATTAATTTAGCAAAATATAAATCAGCAAAATATTTTTC 240
Db 858 GACGGCTATAAGAGTTTCATTTATTAATTTAGCAAAATATAAATCAGCAAAATATTTTTC 917

Qy 241 TTGACTAAGCTTAAACGACGCCGTTAAACATTTTC 274
Db 918 TTGACTAAGCTTAAACGACGCCGTTAAACATTTTC 951

RESULT 6
US-10-800-161-3
; Sequence 3, Application US/10800161
; Publication No. US20040154051A1
; GENERAL INFORMATION:
; APPLICANT: Cade, Rebecca M
; APPLICANT: Dietrich, Robert A
; APPLICANT: Lawton, Kay Ann
; TITLE OF INVENTION: INDUCIBLE PROMOTERS
; FILE REFERENCE: A-31089CIP1
; CURRENT APPLICATION NUMBER: US/10/800,161
; CURRENT FILING DATE: 2004-03-12
; PRIOR APPLICATION NUMBER: 60/171,008
; PRIOR FILING DATE: 1999-12-15
; PRIOR APPLICATION NUMBER: 60/175,519
; PRIOR FILING DATE: 2000-01-11
; NUMBER OF SEQ ID NOS: 31
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3
; LENGTH: 1700

; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (365)..(374)
; OTHER INFORMATION: TCA1 motif
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (426)..(435)
; OTHER INFORMATION: TCA1 motif
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (609)..(614)
; OTHER INFORMATION: MYCATR22 element
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (646)..(665)
; OTHER INFORMATION: CAMV ASI salicylic acid response element
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (707)..(712)
; OTHER INFORMATION: PAL BOX
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (757)..(762)
; OTHER INFORMATION: HEXAMERAT 4 element
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (863)..(1228)
; OTHER INFORMATION: N116 genomic coding region
US-10-800-161-3

Query Match 99.4%; Score 272.4; DB 7; Length 1700;
Best Local Similarity 99.6%; Pred. No. 4.1e-60;
Matches 273; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 TCTAGAAATATAGCGGATACCAATTTTCCACATGAGCTTCCTTTATTCCTCAAAAGTCA 60
Db 578 TCTAAAAATATAGCGGATACCAATTTTCCACATGAGCTTCCTTTATTCCTCAAAAGTCA 637

Qy 61 ATAAAGTGTGACGTCATGATACCTTACGCTTTAAACATCGCATGATGTCATTAGCAT 120
Db 638 ATAAAGTGTGACGTCATGATACCTTACGCTTTAAACATCGCATGATGTCATTAGCAT 697

Qy 121 CAATCTCCACCGTCCAAATTTATTTAGTTGTTGACAATATCGACCGTCTTAAGTTCACAC 180
Db 698 CAATCTCCACCGTCCAAATTTATTTAGTTGTTGACAATATCGACCGTCTTAAGTTCACAC 757

Qy 181 GACGGCTATAAGAGTTTCATTTATTAATTTAGCAAAATATAAATCAGCAAAATATTTTTC 240
Db 758 GACGGCTATAAGAGTTTCATTTATTAATTTAGCAAAATATAAATCAGCAAAATATTTTTC 817

Qy 241 TTGACTAAGCTTAAACGACGCCGTTAAACATTTTC 274
Db 818 TTGACTAAGCTTAAACGACGCCGTTAAACATTTTC 851

RESULT 7
US-10-760-752-3
; Sequence 3, Application US/10760752
; Publication No. US20040248303A1
; GENERAL INFORMATION:
; APPLICANT: Cade, Rebecca M
; APPLICANT: Dietrich, Robert A
; TITLE OF INVENTION: Inducible Promoter Fragment
; FILE REFERENCE: A-31089DIV
; CURRENT APPLICATION NUMBER: US/10/760,752
; CURRENT FILING DATE: 2004-01-20
; PRIOR APPLICATION NUMBER: 09/733685
; PRIOR FILING DATE: 2000-12-8
; PRIOR APPLICATION NUMBER: 60/171,008
; PRIOR FILING DATE: 1999-12-15
; PRIOR APPLICATION NUMBER: 60/175,519
```



```
; PRIOR FILING DATE: 2000-01-11
; NUMBER OF SEQ ID NOS: 23
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3
; LENGTH: 1700
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (365)..(374)
; OTHER INFORMATION: TC1 motif
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (426)..(435)
; OTHER INFORMATION: TC1 motif
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (609)..(614)
; OTHER INFORMATION: MYCATR22 element
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (646)..(665)
; OTHER INFORMATION: CANV AS1 salicylic acid response element
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (707)..(712)
; OTHER INFORMATION: PAL BOX
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (757)..(762)
; OTHER INFORMATION: HEXAMERAT 4 element
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (863)..(1228)
; OTHER INFORMATION: NI16 genomic coding region
US-10-760-752-3
```

```
Query Match 99.4%; Score 272.4; DB 8; Length 1700;
Best Local Similarity 99.6%; Pred. No. 4.1e-60;
Matches 273; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
```

```
QY 1 TCTAGAAATATAGCGATACCAATTTTCCACATGCGATCGATGATGTCATTAGCAT 60
|||
|||
|||
578 TCTAAATAATATAGCGATACCAATTTTCCACATGCGATCGATGATGTCATTAGCAT 637
|||
|||
|||
61 ATAAAGTGAGCGTCATGATCTAGCGCTTTTAAACATCGCATGATGTCATTAGCAT 120
|||
|||
|||
638 ATAAAGTGAGCGTCATGATCTAGCGCTTTTAAACATCGCATGATGTCATTAGCAT 697
|||
|||
|||
121 CAATCTCCACCGTCCCAATTTATTAGTTGTTGACAAATATCGACCGCTTAAGTTCCACACC 180
|||
|||
|||
698 CAATCTCCACCGTCCCAATTTATTAGTTGTTGACAAATATCGACCGCTTAAGTTCCACACC 757
|||
|||
|||
181 GACGGCTATAAGGTTTCATTATTAATTTTACCAATAAATACGCAATAATTTTTC 240
|||
|||
|||
758 GACGGCTATAAGGTTTCATTATTAATTTTACCAATAAATACGCAATAATTTTTC 817
|||
|||
|||
241 TTGACTTAAGCTTAAACGACGCGCTTAACATTTC 274
|||
|||
|||
818 TTGACTTAAGCTTAAACGACGCGCTTAACATTTC 851
|||
|||
|||
```

```
RESULT 8
US-10-800-161-1
; Sequence 1, Application US/10800161
; Publication No. US20040154051A1
; GENERAL INFORMATION:
; APPLICANT: Cade, Rebecca M
; APPLICANT: Dietrich, Robert A
; APPLICANT: Lawton, Kay Ann
; TITLE OF INVENTION: INDUCIBLE PROMOTERS
; FILE REFERENCE: A-31089CIP1
; CURRENT APPLICATION NUMBER: US/10/800,161
```

```
; CURRENT FILING DATE: 2004-03-12
; PRIOR APPLICATION NUMBER: 60/171,008
; PRIOR FILING DATE: 1999-12-15
; PRIOR APPLICATION NUMBER: 60/175,519
; PRIOR FILING DATE: 2000-01-11
; NUMBER OF SEQ ID NOS: 31
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 509
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (68)..(433)
; OTHER INFORMATION: Gene product NI16
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (142)..(147)
; OTHER INFORMATION: Sali site
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (344)..(349)
; OTHER INFORMATION: EcoRI site
US-10-800-161-1
```

```
Query Match 19.3%; Score 52.8; DB 7; Length 509;
Best Local Similarity 96.4%; Pred. No. 0.0015;
Matches 54; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
```

```
QY 219 AAATCAGCAATAATTTTCTTGACTTAAGCTTAAACGACGCCGTTAACATTTC 274
|||
|||
|||
DB 1 AAATCAGCAATAATTTTCTTGACTTAAGCTTAAACGACGCCGTTAACATTTC 56
|||
|||
|||
```

```
RESULT 9
US-10-760-752-1
; Sequence 1, Application US/10760752
; Publication No. US20040248303A1
; GENERAL INFORMATION:
; APPLICANT: Cade, Rebecca M
; APPLICANT: Dietrich, Robert A
; TITLE OF INVENTION: Inducible Promoter Fragment
; FILE REFERENCE: A-31089DIV
; CURRENT APPLICATION NUMBER: US/10/760,752
; CURRENT FILING DATE: 2004-01-20
; PRIOR APPLICATION NUMBER: 09/733685
; PRIOR FILING DATE: 2000-12-8
; PRIOR APPLICATION NUMBER: 60/171,008
; PRIOR FILING DATE: 1999-12-15
; PRIOR APPLICATION NUMBER: 60/175,519
; PRIOR FILING DATE: 2000-01-11
; NUMBER OF SEQ ID NOS: 23
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 509
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (68)..(433)
; OTHER INFORMATION: gene product NI16
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (142)..(147)
; OTHER INFORMATION: Sali site
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (344)..(349)
; OTHER INFORMATION: EcoRI site
US-10-760-752-1

Query Match 19.3%; Score 52.8; DB 8; Length 509;
Best Local Similarity 96.4%; Pred. No. 0.0015;
```

```
Matches 54; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 219 AAAATCAGCAATAATTTTCTTGACTAAGCTTAACGACGCGTTACATTTTC 274
Db 1 AAAATCAGCAATAATTTTCTTGACTAAGCTTAACGACGCGTTACATTTTC 56

RESULT 10
US-10-027-632-102510/c
; Sequence 102510, Application US/10027632
; Publication No. US20020198371A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.129
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 102510
; LENGTH: 2469
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-102510

Query Match 15.8%; Score 43.4; DB 5; Length 2469;
Best Local Similarity 50.2%; Pred. No. 0.74;
Matches 107; Conservative 0; Mismatches 106; Indels 0; Gaps 0;
QY 22 AATTTTCCACATGCACTTCTTTTATTCAAAAGTCAATAAAGTGTGACGTCATGATA 81
Db 2169 AAAATATCTGGAATATATCTTCTTTATTTGAGAATTTTCAAAAATTAAGATCTCAAAAAC 2110
QY 82 CTTACGCTTTAAACATCGCATGATGTCTATTCATCATCAATCTCCACGTCCTCAATTTA 141
Db 2109 AGCATAGTTTGTAGTGAAGAATAAATATCATTTACAGAATGAAATCTTAATGTCAAAAATA 2050
QY 142 TTTAGTTGTTGACAATATCGCGTCTAAAGTTCCACACCGCGCTATAAGAGTTTCATT 201
Db 2049 ATTTCTTAAAGGATTCACAAATGAAACATAATATATATATCCGCTTGAAAAACCTCATT 1990
QY 202 ATAAATTTTAGCAAAATAAATAATCAGCAATAAT 234
Db 1989 TTATATTTGCCAAAATAAAGACAGAGGAAT 1957

RESULT 11
US-10-027-632-102511/c
; Sequence 102511, Application US/10027632
; Publication No. US20020198371A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.129
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 112106
; LENGTH: 2469
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-102511

Query Match 15.8%; Score 43.4; DB 5; Length 2469;
Best Local Similarity 50.2%; Pred. No. 0.74;
Matches 107; Conservative 0; Mismatches 106; Indels 0; Gaps 0;
QY 22 AATTTTCCACATGCACTTCTTTTATTCAAAAGTCAATAAAGTGTGACGTCATGATA 81
Db 2169 AAAATATCTGGAATATATCTTCTTTATTTGAGAATTTTCAAAAATTAAGATCTCAAAAAC 2110
QY 82 CTTACGCTTTAAACATCGCATGATGTCTATTCATCATCAATCTCCACGTCCTCAATTTA 141
Db 2109 AGCATAGTTTGTAGTGAAGAATAAATATCATTTACAGAATGAAATCTTAATGTCAAAAATA 2050
QY 142 TTTAGTTGTTGACAATATCGCGTCTAAAGTTCCACACCGCGCTATAAGAGTTTCATT 201
Db 2049 ATTTCTTAAAGGATTCACAAATGAAACATAATATATATATCCGCTTGAAAAACCTCATT 1990
QY 202 ATAAATTTTAGCAAAATAAATAATCAGCAATAAT 234
Db 1989 TTATATTTGCCAAAATAAAGACAGAGGAAT 1957
```

```
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 102511
; LENGTH: 2469
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-102511

Query Match 15.8%; Score 43.4; DB 5; Length 2469;
Best Local Similarity 50.2%; Pred. No. 0.74;
Matches 107; Conservative 0; Mismatches 106; Indels 0; Gaps 0;
QY 22 AATTTTCCACATGCACTTCTTTTATTCAAAAGTCAATAAAGTGTGACGTCATGATA 81
Db 2169 AAAATATCTGGAATATATCTTCTTTATTTGAGAATTTTCAAAAATTAAGATCTCAAAAAC 2110
QY 82 CTTACGCTTTAAACATCGCATGATGTCTATTCATCATCAATCTCCACGTCCTCAATTTA 141
Db 2109 AGCATAGTTTGTAGTGAAGAATAAATATCATTTACAGAATGAAATCTTAATGTCAAAAATA 2050
QY 142 TTTAGTTGTTGACAATATCGCGTCTAAAGTTCCACACCGCGCTATAAGAGTTTCATT 201
Db 2049 ATTTCTTAAAGGATTCACAAATGAAACATAATATATATATCCGCTTGAAAAACCTCATT 1990
QY 202 ATAAATTTTAGCAAAATAAATAATCAGCAATAAT 234
Db 1989 TTATATTTGCCAAAATAAAGACAGAGGAAT 1957

RESULT 12
US-10-027-632-112106
; Sequence 112106, Application US/10027632
; Publication No. US20020198371A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.129
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 112106
; LENGTH: 2469
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-112106
```





**This Page Blank (uspto)**

Result No.	Score	Query		Length	DB	ID	Description
		Match	%				
C 1	39	14.2	197096	7	US-11-121-086-107	Sequence 107, Appl	
C 2	37.8	13.8	319608	7	US-11-145-703-1	Sequence 1, Appl	
C 3	32.6	11.9	1082144	7	US-11-117-187-211	Sequence 211, Appl	
C 4	32	11.7	1765	6	US-10-750-185-59119	Sequence 59119, A	
C 5	32	11.7	1765	6	US-10-750-623-59119	Sequence 59119, A	
C 6	31.8	11.6	100000	7	US-11-124-368A-2918	Sequence 2918, Ap	
C 7	31.4	11.5	3857	7	US-11-136-527-3203	Sequence 3203, Ap	
C 8	31.2	11.4	5152	6	US-10-240-708-73	Sequence 73, Appl	
C 9	31	11.3	1140	6	US-10-750-185-39222	Sequence 39222, A	
C 10	31	11.3	1140	6	US-10-750-623-39222	Sequence 39222, A	
C 11	31	11.3	1272	6	US-10-750-185-37551	Sequence 37551, A	
C 12	31	11.3	1272	6	US-10-750-623-37551	Sequence 37551, A	
C 13	31	11.3	1400	7	US-11-136-527-6695	Sequence 6695, Ap	
C 14	31	11.3	2217	7	US-11-136-527-2599	Sequence 2599, Ap	
C 15	30.6	11.2	4627	6	US-10-775-169-206	Sequence 206, App	
C 16	30.6	11.2	4627	7	US-11-186-284-52	Sequence 52, Appl	
C 17	30.6	11.2	181172	7	US-11-121-086-41	Sequence 41, Appl	
C 18	30.4	11.1	600	7	US-11-136-527-4185	Sequence 4185, Ap	
C 19	30.4	11.1	600	7	US-11-136-527-7677	Sequence 7677, Ap	
C 20	30.4	11.1	1538	6	US-10-750-185-26322	Sequence 26322, A	
C 21	30.4	11.1	1538	6	US-10-750-623-26322	Sequence 26322, A	
C 22	30.4	11.1	1791	7	US-11-136-527-89	Sequence 89, Appl	
C 23	30.4	11.1	2043	7	US-11-136-527-3581	Sequence 3581, Ap	

```

; APPLICANT: Bougueleret, Lydie
; APPLICANT: Bihain, Bernard
; APPLICANT: Essioux, Laurent
; TITLE OF INVENTION: SCHIZOPHRENIA ASSOCIATED GENES, PROTEINS AND BIALLELIC MARKERS
; FILE REFERENCE: 53 US16.DIV
; CURRENT APPLICATION NUMBER: US/11/145,703
; CURRENT FILING DATE: 2000-03-30
; PRIOR APPLICATION NUMBER: US/10/147,603
; PRIOR FILING DATE: 2000-03-30
; PRIOR APPLICATION NUMBER: 09/539,333
; PRIOR FILING DATE: 2000-03-30
; PRIOR APPLICATION NUMBER: US 60/126,903
; PRIOR FILING DATE: 1999-03-30
; PRIOR APPLICATION NUMBER: US 60/131,971
; PRIOR FILING DATE: 1999-04-30
; PRIOR APPLICATION NUMBER: US 60/132,065
; PRIOR FILING DATE: 1999-04-30
; PRIOR APPLICATION NUMBER: US 60/143,928
; PRIOR FILING DATE: 1999-07-14
; PRIOR APPLICATION NUMBER: US 60/145,915
; PRIOR FILING DATE: 1999-07-27
; PRIOR APPLICATION NUMBER: US 60/146,453
; PRIOR FILING DATE: 1999-07-29
; PRIOR APPLICATION NUMBER: US 60/146,452
; PRIOR FILING DATE: 1999-07-29
; PRIOR APPLICATION NUMBER: US 60/162,288
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 234
; SOFTWARE: Patent.pm
; SEQ ID NO 1
; LENGTH: 319608
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: 31..1107
; OTHER INFORMATION: 5'regulatory region g35018 gene
; FEATURE:
; NAME/KEY: exon
; LOCATION: 1108..1289
; OTHER INFORMATION: exon A g35018 gene
; FEATURE:
; NAME/KEY: exon
; LOCATION: 14877..14920
; OTHER INFORMATION: exon B g35018 gene
; FEATURE:
; NAME/KEY: exon
; LOCATION: 18778..18862
; OTHER INFORMATION: exon Bbis g35018 gene
; FEATURE:
; NAME/KEY: exon
; LOCATION: 25593..25740
; OTHER INFORMATION: exon C g35018 gene
; FEATURE:
; NAME/KEY: exon
; LOCATION: 29388..29502
; OTHER INFORMATION: exon D g35018 gene
; FEATURE:
; NAME/KEY: exon
; LOCATION: 29967..30282
; OTHER INFORMATION: exon E g35018 gene
; FEATURE:
; NAME/KEY: exon
; LOCATION: 64666..64812
; OTHER INFORMATION: exon F g35018 gene
; FEATURE:
; NAME/KEY: exon
; LOCATION: 65505..65853
; OTHER INFORMATION: exon G g35018 gene
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: 65854..67854
; OTHER INFORMATION: exon Nbis complement g34872 gene

; OTHER INFORMATION: 3'regulatory region g35018 gene
; FEATURE:
; NAME/KEY: exon
; LOCATION: 94124..94964
; OTHER INFORMATION: exon g35017
; FEATURE:
; NAME/KEY: exon
; LOCATION: 201188..201234
; OTHER INFORMATION: exon S g35030 gene
; FEATURE:
; NAME/KEY: exon
; LOCATION: 214676..214793
; OTHER INFORMATION: exon T g35030 gene
; FEATURE:
; NAME/KEY: exon
; LOCATION: 215702..215746
; OTHER INFORMATION: exon U g35030 gene
; FEATURE:
; NAME/KEY: exon
; LOCATION: 216836..216915
; OTHER INFORMATION: exon V g35030 gene
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: 213818..215818
; OTHER INFORMATION: 3'regulatory region g34872 gene
; FEATURE:
; NAME/KEY: exon
; LOCATION: 215819..215941
; OTHER INFORMATION: exon R complement g34872 gene
; FEATURE:
; NAME/KEY: exon
; LOCATION: 215819..215975
; OTHER INFORMATION: exon Rbis complement g34872 gene
; FEATURE:
; NAME/KEY: exon
; LOCATION: 216661..216952
; OTHER INFORMATION: exon Qbis complement g34872 gene
; FEATURE:
; NAME/KEY: exon
; LOCATION: 216661..217061
; OTHER INFORMATION: exon Q complement g34872 gene
; FEATURE:
; NAME/KEY: exon
; LOCATION: 217027..217061
; OTHER INFORMATION: exon Q1 complement g34872 gene
; FEATURE:
; NAME/KEY: exon
; LOCATION: 229647..229742
; OTHER INFORMATION: exon X complement g34872 gene
; FEATURE:
; NAME/KEY: exon
; LOCATION: 230408..230721
; OTHER INFORMATION: exon P complement g34872 gene
; FEATURE:
; NAME/KEY: exon
; LOCATION: 231272..231412
; OTHER INFORMATION: exon Obis complement g34872 gene
; FEATURE:
; NAME/KEY: exon
; LOCATION: 231787..231880
; OTHER INFORMATION: exon O2 complement g34872 gene
; FEATURE:
; NAME/KEY: exon
; LOCATION: 231870..231879
; OTHER INFORMATION: exon O1 complement g34872 gene
; FEATURE:
; NAME/KEY: exon
; LOCATION: 234174..234321
; OTHER INFORMATION: exon O complement g34872 gene
; FEATURE:
; NAME/KEY: exon
; LOCATION: 237406..237428
; OTHER INFORMATION: exon Nbis complement g34872 gene
```



```

; TITLE OF INVENTION: PLANT ARTIFICIAL CHROMOSOME COMPOSITIONS AND METHODS
; FILE REFERENCE: ARCD:309US
; CURRENT APPLICATION NUMBER: US/11/117,187
; CURRENT FILING DATE: 2005-04-28
; PRIOR APPLICATION NUMBER: US/09/531,120
; PRIOR FILING DATE: 2000-03-17
; PRIOR APPLICATION NUMBER: 60/125,219
; PRIOR FILING DATE: 1999-03-18
; NUMBER OF SEQ ID NOS: 212
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 211
; LENGTH: 1082144
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
; US-11-117-187-211

Query Match      11.9%; Score 32.6; DB 7; Length 1082144;
Best Local Similarity 50.3%; Pred. No. 1e+02;
Matches 80; Conservative 0; Mismatches 79; Indels 0; Gaps 0;

QY   81 ACTTAGCGTTTAAACATCGCATGATGTCATTAGCATTCAATCCACCGCTCAATTT 140
Db   345257 ATTGGTGCTTAACAATTCCTAAACTTTTATGATCCTATTCAATCTAAACCGTTCAATAA 345198

QY   141 ATTTAGTTGTGACATATATCGCGCTGAAGTTCCACACCGCGCTATAAGAGTTTCAT 200
Db   345197 AATTATTTCCCTTCAATATAAACCGTTCAACAAAATAAGGAATCTATGTCGTTTTTGAA 345138

QY   201 TATAAATTTTTCAGCAAAATAAAATTCAGCAAAATAATTTTTT 239
Db   345137 ATCAAATAATATCACATTTTCATCCCTTAAAAACACATATAT 345099

RESULT 4
US-10-750-185-59119
; Sequence 59119, Application US/10750185
; Publication No. US20050260603A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: FANTIN, Dennis
; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MM1100-2
; CURRENT APPLICATION NUMBER: US/10/750,185
; CURRENT FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; PRIOR FILING DATE: 2002-12-31
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: Patentin version 3.1
; SEQ ID NO 59119
; LENGTH: 1765
; TYPE: DNA
; ORGANISM: Bovine 19866880452391
; US-10-750-185-59119

Query Match      11.7%; Score 32; DB 6; Length 1765;
Best Local Similarity 56.7%; Pred. No. 20;
Matches 59; Conservative 0; Mismatches 45; Indels 0; Gaps 0;

QY   145 AGTTGTTGACAATATCGACCGTCTTAAGTTCCACACCGCGCTATAAGAGTTTCATTATA 204
Db   .936 AGTTGAAAACCTCTGTCCACACAGAAATCTGCACACAGAGGTTTACAGCAGTTTAGTCATA 995

QY   205 AATTTTAGCAAAATAAAATCAGCAAAATAATTTTTTCTTGACTAA 248
Db   996 ATTGCCAAAATTGGAAGCCCAAGAATATTTCTTTGTTAGTAA 1039

RESULT 5

```









**This Page Blank (uspto)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2006 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 12, 2006, 00:16:10 ; Search time 148 Seconds  
(without alignments)  
3290.891 Million cell updates/sec

Title: US-10-800-161-28

Perfect score: 274

Sequence: 1 tctagaatatccgatac.....acgaacgcgttaacattttc 274

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 1303057 seqs, 888780828 residues

Total number of hits satisfying chosen parameters: 2606114

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- 1: Issued Patents NA:\*
- 2: /cgn2\_6/ptodata/1/ina/1 COMB.seq:\*
- 3: /cgn2\_6/ptodata/1/ina/5 COMB.seq:\*
- 4: /cgn2\_6/ptodata/1/ina/6A COMB.seq:\*
- 5: /cgn2\_6/ptodata/1/ina/6B COMB.seq:\*
- 6: /cgn2\_6/ptodata/1/ina/H COMB.seq:\*
- 7: /cgn2\_6/ptodata/1/ina/PCTUS COMB.seq:\*
- 8: /cgn2\_6/ptodata/1/ina/PP COMB.seq:\*
- 9: /cgn2\_6/ptodata/1/ina/RE COMB.seq:\*
- 9: /cgn2\_6/ptodata/1/ina/backfiles1.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	272.4	99.4	1700	3	US-09-733-685-3
2	52.8	19.3	509	3	US-09-733-685-1
3	39	14.2	601	3	US-09-949-016-83537
4	39	14.2	49440	3	US-09-949-016-14150
5	37.8	13.8	319608	3	US-09-539-333D-1
6	37.8	13.8	319608	3	US-09-679-409-1
7	37.2	13.6	1852	3	US-09-813-356-2
8	37.2	13.6	1963	3	US-09-813-356-1
9	37.2	13.6	1963	3	US-09-813-356-3
10	35.4	12.9	30002	3	US-09-949-016-15866
11	34.6	12.6	1141	3	US-09-806-708B-22
12	34.4	12.6	227390	3	US-09-949-016-12201
13	34.4	12.6	227391	3	US-09-949-016-13365
14	34.2	12.5	601	3	US-09-949-016-112718
15	34.2	12.5	601	3	US-09-949-016-112802
16	34.2	12.5	601	3	US-09-949-016-112886
17	34.2	12.5	601	3	US-09-949-016-112962
18	34.2	12.5	601	3	US-09-949-016-113048
19	34.2	12.5	127280	3	US-09-949-016-14857
20	34.2	12.5	132266	3	US-09-949-016-14860
21	34.2	12.5	150833	3	US-09-949-016-14859
22	34.2	12.5	159963	3	US-09-949-016-14858
23	34.2	12.5	171130	3	US-09-949-016-14861
24	34	12.4	2275	2	US-08-743-637B-2

25	34	12.4	2275	3	US-08-526-840B-2	Sequence 2, Appli
26	33.6	12.3	474	3	US-09-621-976-18033	Sequence 18033, A
27	33.6	12.3	3244	3	US-09-949-016-2708	Sequence 2708, Ap
28	33.6	12.3	3953	3	US-09-220-132-46	Sequence 46, Appl
29	33.6	12.3	3953	3	US-09-949-016-263	Sequence 263, App
30	33.6	12.3	12797	3	US-09-949-016-15274	Sequence 15274, A
31	33.6	12.3	832	3	US-09-621-976-2813	Sequence 2813, Ap
32	33.4	12.2	27363	3	US-09-949-016-17535	Sequence 17535, A
33	33.4	12.2	67755	3	US-09-949-016-13703	Sequence 13703, A
34	33.2	12.1	466	2	US-08-743-637B-29	Sequence 29, Appl
35	33.2	12.1	466	3	US-08-526-840B-29	Sequence 29, Appl
36	33.2	12.1	600	3	US-09-540-236-1274	Sequence 1274, Ap
37	33.2	12.1	92407	3	US-09-596-002-36	Sequence 36, Appl
38	33.2	12.1	139049	3	US-09-949-016-17030	Sequence 17030, A
39	33.2	12.1	139036	3	US-09-949-016-11782	Sequence 11782, A
40	33.2	12.1	139952	3	US-09-949-016-13280	Sequence 13280, A
41	33	12.0	480	3	US-09-621-976-13994	Sequence 13994, A
42	33	12.0	10320	3	US-09-949-016-11778	Sequence 11778, A
43	32.8	12.0	198	3	US-09-513-999C-19031	Sequence 19031, A
44	32.8	12.0	893	3	US-09-270-767-15009	Sequence 15009, A
45	32.8	12.0	1664976	3	US-08-916-421B-1	Sequence 1, Appli

## ALIGNMENTS

RESULT 1  
US-09-733-685-3  
; Sequence 3, Application US/09733685  
; Patent No. 6706952  
; GENERAL INFORMATION:  
; APPLICANT: Cade, Rebecca M  
; APPLICANT: Dietrich, Robert A  
; TITLE OF INVENTION: GENES ENCODING PROTEINS INVOLVED IN THE REGULATION OF  
; FILE REFERENCE: A-31089A  
; CURRENT APPLICATION NUMBER: US/09/733,685  
; PRIOR FILING DATE: 2000-12-08  
; PRIOR APPLICATION NUMBER: 60/171,008  
; PRIOR FILING DATE: 1999-12-15  
; PRIOR APPLICATION NUMBER: 60/175,519  
; PRIOR FILING DATE: 2000-01-11  
; NUMBER OF SEQ ID NOS: 23  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 3  
; LENGTH: 1700  
; TYPE: DNA  
; ORGANISM: Arabidopsis thaliana  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (365)..(374)  
; OTHER INFORMATION: TCAL motif  
; NAME/KEY: misc feature  
; LOCATION: (426)..(435)  
; OTHER INFORMATION: TCAL motif  
; NAME/KEY: misc feature  
; LOCATION: (609)..(614)  
; OTHER INFORMATION: MYCATR22 element  
; NAME/KEY: misc feature  
; LOCATION: (646)..(665)  
; OTHER INFORMATION: CAMV ASI salicylic acid response element  
; NAME/KEY: misc feature  
; LOCATION: (707)..(712)  
; OTHER INFORMATION: PAL BOX  
; NAME/KEY: misc feature  
; LOCATION: (757)..(762)  
; OTHER INFORMATION: HEXAMERAT 4 element  
; NAME/KEY: misc feature  
; LOCATION: (863)..(1228)  
; OTHER INFORMATION: N116 genomic coding region  
US-09-733-685-3

Query Match 99.4%; Score 272.4; DB 3; Length 1700;





```
QY      189 TAAGATTCATTATAATTTTAGCAAAATAAAATCAGCAATAATTTTCTTG 243
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      24627 TAACATGTTTAATAGATGACCAATAATGATTTTAATAGTCCTTCTTG 24681

RESULT 5
US-09-539-333D-1
; Sequence 1, Application US/09539333D
; Patent No. 6476208
; GENERAL INFORMATION:
; APPLICANT: Cohen, Daniel
; APPLICANT: Blumenfeld, Marta
; APPLICANT: Chumakov, Ilya
; APPLICANT: Bougueleret, Lydie
; APPLICANT: Bihain, Bernard
; APPLICANT: Essioux, Laurent
; TITLE OF INVENTION: SCHIZOPHRENIA ASSOCIATED GENES, PROTEINS AND BIALLELIC MARKERS
; FILE REFERENCE: GENSET.047AUS
; CURRENT APPLICATION NUMBER: US/09/539,333D
; CURRENT FILING DATE: 2000-03-30
; PRIOR APPLICATION NUMBER: US 60/126,903
; PRIOR FILING DATE: 1999-03-30
; PRIOR APPLICATION NUMBER: US 60/131,971
; PRIOR FILING DATE: 1999-04-30
; PRIOR APPLICATION NUMBER: US 60/132,065
; PRIOR FILING DATE: 1999-04-30
; PRIOR APPLICATION NUMBER: US 60/143,928
; PRIOR FILING DATE: 1999-07-14
; PRIOR APPLICATION NUMBER: US 60/145,915
; PRIOR FILING DATE: 1999-07-27
; PRIOR APPLICATION NUMBER: US 60/146,453
; PRIOR FILING DATE: 1999-07-29
; PRIOR APPLICATION NUMBER: US 60/146,452
; PRIOR FILING DATE: 1999-07-29
; PRIOR APPLICATION NUMBER: US 60/162,288
; PRIOR FILING DATE: 1999-10-28
; PRIOR APPLICATION NUMBER: US 09/416,384
; PRIOR FILING DATE: 1999-10-12
; NUMBER OF SEQ ID NOS: 231
; SOFTWARE: Patent.pm
; SEQ ID NO 1
; LENGTH: 319608
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 31..1107
; OTHER INFORMATION: 5'regulatory region g35018 gene
; FEATURE:
; NAME/KEY: exon
; LOCATION: 1108..1289
; OTHER INFORMATION: exon A g35018 gene
; FEATURE:
; NAME/KEY: exon
; LOCATION: 14877..14920
; OTHER INFORMATION: exon B g35018 gene
; FEATURE:
; NAME/KEY: exon
; LOCATION: 18778..18862
; OTHER INFORMATION: exon Bbis g35018 gene
; FEATURE:
; NAME/KEY: exon
; LOCATION: 25593..25740
; OTHER INFORMATION: exon C g35018 gene
; FEATURE:
; NAME/KEY: exon
; LOCATION: 29388..29502
; OTHER INFORMATION: exon D g35018 gene
; FEATURE:
; NAME/KEY: exon
; LOCATION: 29967..30282
; OTHER INFORMATION: exon E g35018 gene
; FEATURE:
; NAME/KEY: exon
; LOCATION: 64666..64812
; OTHER INFORMATION: exon F g35018 gene
; FEATURE:
; NAME/KEY: exon
; LOCATION: 65505..65853
; OTHER INFORMATION: exon G g35018 gene
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 65854..67854
; OTHER INFORMATION: 3'regulatory region g35018 gene
; FEATURE:
; NAME/KEY: exon
; LOCATION: 94124..94964
; OTHER INFORMATION: exon g35017
; FEATURE:
; NAME/KEY: exon
; LOCATION: 201188..201234
; OTHER INFORMATION: exon S g35030 gene
; FEATURE:
; NAME/KEY: exon
; LOCATION: 214676..214793
; OTHER INFORMATION: exon T g35030 gene
; FEATURE:
; NAME/KEY: exon
; LOCATION: 215702..215746
; OTHER INFORMATION: exon U g35030 gene
; FEATURE:
; NAME/KEY: exon
; LOCATION: 216836..216915
; OTHER INFORMATION: exon V g35030 gene
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 213818..215818
; OTHER INFORMATION: 3'regulatory region g34872 gene
; FEATURE:
; NAME/KEY: exon
; LOCATION: 215819..215941
; OTHER INFORMATION: exon R complement g34872 gene
; FEATURE:
; NAME/KEY: exon
; LOCATION: 215819..215975
; OTHER INFORMATION: exon Rbis complement g34872 gene
; FEATURE:
; NAME/KEY: exon
; LOCATION: 216661..216952
; OTHER INFORMATION: exon Qbis complement g34872 gene
; FEATURE:
; NAME/KEY: exon
; LOCATION: 216661..217061
; OTHER INFORMATION: exon Q complement g34872 gene
; FEATURE:
; NAME/KEY: exon
; LOCATION: 217027..217061
; OTHER INFORMATION: exon Q1 complement g34872 gene
; FEATURE:
; NAME/KEY: exon
; LOCATION: 229647..229742
; OTHER INFORMATION: exon X complement g34872 gene
; FEATURE:
; NAME/KEY: exon
; LOCATION: 230408..230721
; OTHER INFORMATION: exon P complement g34872 gene
; FEATURE:
; NAME/KEY: exon
; LOCATION: 231272..231412
; OTHER INFORMATION: exon Obis complement g34872 gene
; FEATURE:
; NAME/KEY: exon
; LOCATION: 231787..231880
; OTHER INFORMATION: exon O2 complement g34872 gene
; FEATURE:
; NAME/KEY: exon
```

LOCATION: 231870..231879  
OTHER INFORMATION: exon O1 complement g34872 gene  
FEATURE:  
NAME/KEY: exon  
LOCATION: 234174..234321  
OTHER INFORMATION: exon O complement g34872 gene  
FEATURE:  
NAME/KEY: exon  
LOCATION: 237406..237428  
OTHER INFORMATION: exon Nbis complement g34872 gene  
FEATURE:  
NAME/KEY: exon  
LOCATION: 239719..239807  
OTHER INFORMATION: exon N2 complement g34872 gene  
FEATURE:  
NAME/KEY: exon  
LOCATION: 239719..239853  
OTHER INFORMATION: exon N complement g34872 gene  
FEATURE:  
NAME/KEY: exon  
LOCATION: 240528..240569  
OTHER INFORMATION: exon M1117 complement g34872 gene  
FEATURE:  
NAME/KEY: exon  
LOCATION: 240528..240596  
OTHER INFORMATION: exon M1090 complement g34872 gene  
FEATURE:  
NAME/KEY: exon  
LOCATION: 240528..240617  
OTHER INFORMATION: exon M1069 complement g34872 gene  
FEATURE:  
NAME/KEY: exon  
LOCATION: 240528..240644  
OTHER INFORMATION: exon MS2 complement g34872 gene  
FEATURE:  
NAME/KEY: exon  
LOCATION: 240528..240824  
OTHER INFORMATION: exon M862 complement g34872 gene  
FEATURE:  
NAME/KEY: exon  
LOCATION: 240528..240994  
OTHER INFORMATION: exon M692 complement g34872 gene  
FEATURE:  
NAME/KEY: exon  
LOCATION: 240528..241685  
OTHER INFORMATION: exon M1 complement g34872 gene  
FEATURE:  
NAME/KEY: exon  
LOCATION: 240800..240993  
OTHER INFORMATION: exon MS1 complement g34872 gene  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: 241686..243685  
OTHER INFORMATION: 5'regulatory region g34872 gene  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: 290652..292652  
OTHER INFORMATION: 3'regulatory region g34665 gene  
FEATURE:  
NAME/KEY: exon  
LOCATION: 292653..292841

Query Match 13.8%; Score 37.8; DB 3; Length 319608;  
Best Local Similarity 52.9%; Pred. No. 1.1;  
Matches 81; Conservative 0; Mismatches 72; Indels 0; Gaps 0;  
QY 38 GACTTCCTTATCCAAAGTCAATAAGTGTGACGTCTACTTACGCTTTAAACA 97  
DB 57795 GAACCTGTATCTCCACATGTGTACATGGTGACTTCCTAATTCACCCATATATACAG 57854  
QY 98 TCGCATGATGATGATAGCATCAATCTCCACCGTCCCAATTTATTAGTTGTGACAAT 157  
DB 57855 TTGCTTGTGAATGCTTTATCTCAATGTACGCGTCCAAAGAGTAAAGAGAAAC 57914

QY 158 ATCGACCGTCTTAAGTTCCACACCGACCGGCTATA 190  
DB 57915 AAAGGAGAAAAATGAGGCACACCCCTCTTTA 57947  
RESULT 6  
US-09-679-409-1  
; Sequence 1, Application US/09679409  
; Patent No. 6555316  
; GENERAL INFORMATION:  
; APPLICANT: Cohen, Daniel  
; APPLICANT: Blumenfeld, Marta  
; APPLICANT: Chumakov, Ilya  
; APPLICANT: Bougueleret, Lydie  
; APPLICANT: Essioux, Laurent  
; TITLE OF INVENTION: SCHIZOPHRENIA ASSOCIATED GENE, PROTEINS AND BIALLELIC MARKERS  
; FILE REFERENCE: 53 US15 CIP  
; CURRENT APPLICATION NUMBER: US/09/679,409  
; CURRENT FILING DATE: 2000-10-03  
; PRIOR APPLICATION NUMBER: 09/539,333  
; PRIOR FILING DATE: 2000-03-03  
; PRIOR APPLICATION NUMBER: 09/416,384  
; PRIOR FILING DATE: 1999-10-12  
; PRIOR APPLICATION NUMBER: 60/168,088  
; PRIOR FILING DATE: 1999-11-30  
; NUMBER OF SEQ ID NOS: 134  
; SOFTWARE: Patent.pm  
; SEQ ID NO 1  
; LENGTH: 319608  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: 199122..201122  
; OTHER INFORMATION: 5'regulatory region  
; NAME/KEY: exon  
; LOCATION: 201123..201234  
; OTHER INFORMATION: exon S  
; NAME/KEY: exon  
; LOCATION: 201123..201560  
; OTHER INFORMATION: exon S2  
; NAME/KEY: exon  
; LOCATION: 214676..214793  
; OTHER INFORMATION: exon T  
; NAME/KEY: exon  
; LOCATION: 215702..215746  
; OTHER INFORMATION: exon U  
; NAME/KEY: exon  
; LOCATION: 216836..216994  
; OTHER INFORMATION: exon V  
; NAME/KEY: exon  
; LOCATION: 216836..217077  
; OTHER INFORMATION: exon V2  
; NAME/KEY: exon  
; LOCATION: 217671..217764  
; OTHER INFORMATION: exon V1  
; NAME/KEY: exon  
; LOCATION: 227655..227736  
; OTHER INFORMATION: exon V4  
; NAME/KEY: exon  
; LOCATION: 238715..238919  
; OTHER INFORMATION: exon V3  
; NAME/KEY: exon  
; LOCATION: 240440..240673  
; OTHER INFORMATION: exon W  
; NAME/KEY: exon  
; LOCATION: 240440..241153  
; OTHER INFORMATION: exon W2  
; NAME/KEY: exon  
; LOCATION: 241072..241291  
; OTHER INFORMATION: exon X  
; NAME/KEY: exon

```
LOCATION: 244353..244561
OTHER INFORMATION: exon Y
NAME/KEY: exon
LOCATION: 246273..247802
OTHER INFORMATION: exon Z
NAME/KEY: misc feature
LOCATION: 247803..249803
OTHER INFORMATION: 3'regulatory region
NAME/KEY: allele
LOCATION: 8316
OTHER INFORMATION: 99-27943-150 : polymorphic base G or C
NAME/KEY: allele
LOCATION: 21672
OTHER INFORMATION: 99-27935-193 : polymorphic base G or C
NAME/KEY: allele
LOCATION: 65485
OTHER INFORMATION: 8-128-33 : polymorphic base C or T
NAME/KEY: allele
LOCATION: 95396
OTHER INFORMATION: 99-31960-363 : polymorphic base A or G
NAME/KEY: allele
LOCATION: 107281
OTHER INFORMATION: 99-24656-260 : polymorphic base A or G
NAME/KEY: allele
LOCATION: 160640
OTHER INFORMATION: 99-24639-163 : polymorphic base A or C
NAME/KEY: allele
LOCATION: 160876
OTHER INFORMATION: 99-24634-108 : polymorphic base A or T
NAME/KEY: allele
LOCATION: 168974
OTHER INFORMATION: 99-7652-162 : polymorphic base A or G
NAME/KEY: allele
LOCATION: 170810
OTHER INFORMATION: 99-16100-147 : polymorphic base A or G
NAME/KEY: allele
LOCATION: 173358
OTHER INFORMATION: 99-5862-167 : polymorphic base A or G
NAME/KEY: allele
LOCATION: 189957
OTHER INFORMATION: 99-5919-215 : polymorphic base A or G
NAME/KEY: allele
LOCATION: 197163
OTHER INFORMATION: 99-24658-410 : polymorphic base A or G
NAME/KEY: allele
LOCATION: 200778
OTHER INFORMATION: 8-303-235 : polymorphic base A or G
NAME/KEY: allele
LOCATION: 202651
OTHER INFORMATION: 8-300-221 : polymorphic base A or G
NAME/KEY: allele
LOCATION: 202679
OTHER INFORMATION: 8-300-193 : polymorphic base A or G
NAME/KEY: allele
LOCATION: 203378
OTHER INFORMATION: 8-299-128 : polymorphic base A or T
NAME/KEY: allele
LOCATION: 204138
OTHER INFORMATION: 8-296-213 : polymorphic base A or T
NAME/KEY: allele
LOCATION: 204605
OTHER INFORMATION: 8-252-190 : polymorphic base C or T
NAME/KEY: allele
LOCATION: 204934
OTHER INFORMATION: 99-24644-194 : polymorphic base A or G
NAME/KEY: allele
LOCATION: 205206
OTHER INFORMATION: 8-295-248 : polymorphic base A or C
NAME/KEY: allele
LOCATION: 205329
OTHER INFORMATION: 8-295-125 : polymorphic base C or T
NAME/KEY: allele
LOCATION: 206064

OTHER INFORMATION: 8-293-130 : polymorphic base A or G
NAME/KEY: allele
LOCATION: 206545
OTHER INFORMATION: 8-292-198 : polymorphic base A or G
NAME/KEY: allele
LOCATION: 207313
OTHER INFORMATION: 8-251-322 : polymorphic base A or G
NAME/KEY: allele
LOCATION: 208285
OTHER INFORMATION: 8-289-322 : polymorphic base A or G
NAME/KEY: allele
LOCATION: 208960
OTHER INFORMATION: 8-287-249 : polymorphic base C or T
NAME/KEY: allele
LOCATION: 209123
OTHER INFORMATION: 8-287-86 : polymorphic base A or T
NAME/KEY: allele
LOCATION: 209631
OTHER INFORMATION: 8-285-319 : polymorphic base A or G
NAME/KEY: allele
LOCATION: 210361
OTHER INFORMATION: 8-283-278 : polymorphic base G or C
NAME/KEY: allele
LOCATION: 210463
OTHER INFORMATION: 8-283-176 : polymorphic base A or G
NAME/KEY: allele
LOCATION: 210486
OTHER INFORMATION: 8-283-153 : polymorphic base G or C
NAME/KEY: allele
LOCATION: 210583
OTHER INFORMATION: 8-283-56 : polymorphic base C or T
NAME/KEY: allele
LOCATION: 210879
OTHER INFORMATION: 8-282-345 : polymorphic base G or C
NAME/KEY: allele
LOCATION: 210964
OTHER INFORMATION: 8-282-260 : polymorphic base G or T
NAME/KEY: allele
LOCATION: 210979
OTHER INFORMATION: 8-282-245 : polymorphic base A or C
NAME/KEY: allele
LOCATION: 211050
OTHER INFORMATION: 8-282-174 : variable motif AAAGG or GAAGGAAGGAAGGAAGGAAGA
NAME/KEY: allele
LOCATION: 211132
OTHER INFORMATION: 8-282-92 : polymorphic base A or T
NAME/KEY: allele
LOCATION: 211247
OTHER INFORMATION: 8-281-367 : polymorphic base A or G
NAME/KEY: allele
LOCATION: 211315
OTHER INFORMATION: 8-281-299 : polymorphic base A or G
NAME/KEY: allele
LOCATION: 211366
OTHER INFORMATION: 8-281-248 : polymorphic base G or C
NAME/KEY: allele
LOCATION: 212520
OTHER INFORMATION: 8-279-197 : polymorphic base A or C
NAME/KEY: allele
LOCATION: 212821
OTHER INFORMATION: 8-278-289 : polymorphic base C or T
NAME/KEY: allele

Query Match 13.8%; Score 37.8; DB 3; Length 319608;
Best Local Similarity 52.9%; Pred. No. 1.1;
Matches 381; Conservative 0; Mismatches 72; Indels 0; Gaps 0;

Qy 38 GACTTCCTTTATCCAAAGTCAATAAAGTGACGTCATGATACCTTACCGCTTTAAACA 97
Db 57795 GAACCTTGATCTCTCCCACTGTGTCTACATGCTGACTTCTTAATTCACCATATATACAG 57854
Qy 98 TCGCATGATGATGTCATTAGCATCAATCTCCACCGTCCAAATTTATTTAGTTGTGCAAT 157
```





; Sequence 13365, Application US/09949016  
; Patent No. 6812339  
; GENERAL INFORMATION:  
; APPLICANT: VENTER, J. Craig et al.  
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF  
; FILE REFERENCE: CL001307  
; CURRENT APPLICATION NUMBER: US/09/949,016  
; CURRENT FILING DATE: 2000-04-14  
; PRIOR APPLICATION NUMBER: 60/241,755  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/237,768  
; PRIOR FILING DATE: 2000-10-03  
; PRIOR APPLICATION NUMBER: 60/231,498  
; PRIOR FILING DATE: 2000-09-08  
; NUMBER OF SEQ ID NOS: 207012  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 13365  
; LENGTH: 227391  
; TYPE: DNA  
; ORGANISM: Human  
; FEATURE:  
; NAME/KEY: misc.feature  
; LOCATION: (1)...((227391))  
; OTHER INFORMATION: n = A,T,C or G  
US-09-949-016-13365

Query Match 12.6%; Score 34.4; DB 3; Length 227391;  
Best Local Similarity 52.0%; Pred. No. 9.3;  
Matches 77; Conservative 0; Mismatches 71; Indels 0; Gaps 0;  
QY 83 TTACGCTTTAAACATCGCATGATGTCATTAGCATCAATCTCCACGGTCAATTAT 142  
DB 161269 TTTCAGAAAAAATAAATGTGTATCAATTACTATTACTTGCATATCTCTTATCT 161210  
QY 143 TTAGTTGTGCAATATCGACGCTTAAGTTCACACCGCGCTAAGAGTTTCATTA 202  
DB 161209 AATTTGTCTCCAAATGTCTCTGGAGAACCCACAGGAGGAAGACTCTCTTTT 161150  
QY 203 TAAATTTTAGCAAAATAAATCAGCAA 230  
DB 161149 TTCCTTAAGTTACATCCATCAGCAA 161122

RESULT 14  
US-09-949-016-112718/c  
; Sequence 112718, Application US/09949016  
; Patent No. 6812339  
; GENERAL INFORMATION:  
; APPLICANT: VENTER, J. Craig et al.  
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF  
; FILE REFERENCE: CL001307  
; CURRENT APPLICATION NUMBER: US/09/949,016  
; CURRENT FILING DATE: 2000-04-14  
; PRIOR APPLICATION NUMBER: 60/241,755  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/237,768  
; PRIOR FILING DATE: 2000-10-03  
; PRIOR APPLICATION NUMBER: 60/231,498  
; PRIOR FILING DATE: 2000-09-08  
; NUMBER OF SEQ ID NOS: 207012  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 112718  
; LENGTH: 601  
; TYPE: DNA  
; ORGANISM: Human  
US-09-949-016-112718

Query Match 12.5%; Score 34.2; DB 3; Length 601;  
Best Local Similarity 48.2%; Pred. No. 2.5;  
Matches 96; Conservative 0; Mismatches 103; Indels 0; Gaps 0;

QY 12 AGCGATACCAATTTTCCACACATGGACTTCCTTTATTCCAAAGTCAATAAAGTGTA 71  
DB 210 AGGCAACCAAGTTTGTCTTTTCTGAATATTTTATTAAGGAATTTATAAATAGGTA 151  
QY 72 CGTCATGATACCTTACGCTTTTAAACATCGCATGATGTCATTAGCATCAATCTCCACC 131  
DB 150 GTCCTTTATATGCTCTTTTATATGACATAGCATATGTCAAACTCATCTTTTGACCACACC 91  
QY 132 GTCCAAATTTATTTAGTTGTCACAAATATCGACCGCTTAAGTTCCACACGACGGCTATAA 191  
DB 90 CTCAAATATGTTATGTCCTCCAACTTCCTTACCTGGTAAGTGGCACTTCTATTCCAGCCA 31  
QY 192 GAGTTTCATTATAAATTTT 210  
DB 30 GAATTCAGAAAGTAATTTT 12

RESULT 15  
US-09-949-016-112802/c  
; Sequence 112802, Application US/09949016  
; Patent No. 6812339  
; GENERAL INFORMATION:  
; APPLICANT: VENTER, J. Craig et al.  
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF  
; FILE REFERENCE: CL001307  
; CURRENT APPLICATION NUMBER: US/09/949,016  
; CURRENT FILING DATE: 2000-04-14  
; PRIOR APPLICATION NUMBER: 60/241,755  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/237,768  
; PRIOR FILING DATE: 2000-10-03  
; PRIOR APPLICATION NUMBER: 60/231,498  
; PRIOR FILING DATE: 2000-09-08  
; NUMBER OF SEQ ID NOS: 207012  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 112802  
; LENGTH: 601  
; TYPE: DNA  
; ORGANISM: Human  
US-09-949-016-112802

Query Match 12.5%; Score 34.2; DB 3; Length 601;  
Best Local Similarity 48.2%; Pred. No. 2.5;  
Matches 96; Conservative 0; Mismatches 103; Indels 0; Gaps 0;

QY 12 AGCGATACCAATTTTCCACACATGGACTTCCTTTATTCCAAAGTCAATAAAGTGTA 71  
DB 210 AGGCAACCAAGTTTGTCTTTTCTGAATATTTTATTAAGGAATTTATAAATAGGTA 151  
QY 72 CGTCATGATACCTTACGCTTTTAAACATCGCATGATGTCATTAGCATCAATCTCCACC 131  
DB 150 GTCCTTTATATGCTCTTTTATATGACATAGCATATGTCAAACTCATCTTTTGACCACACC 91  
QY 132 GTCCAAATTTATTTAGTTGTCACAAATATCGACCGCTTAAGTTCCACACGACGGCTATAA 191  
DB 90 CTCAAATATGTTATGTCCTCCAACTTCCTTACCTGGTAAGTGGCACTTCTATTCCAGCCA 31  
QY 192 GAGTTTCATTATAAATTTT 210  
DB 30 GAATTCAGAAAGTAATTTT 12

Search completed: January 12, 2006, 02:17:53  
Job time : 155 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2006 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 12, 2006, 00:02:06 ; Search time 3802 Seconds  
(without alignments)

3371.820 Million cell updates/sec

Title: US-10-800-161-28

Perfect score: 274

Sequence: 1 tctagaataatgccgatac.....acgacgcgttaacatttc 274

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 41078325 seqs, 23393541228 residues

Total number of hits satisfying chosen parameters: 82156650

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:\*

1: gb\_est1.\*

2: gb\_est2.\*

3: gb\_est3.\*

4: gb\_est4.\*

5: gb\_est5.\*

6: gb\_est6.\*

7: gb\_est7.\*

8: gb\_est8.\*

9: gb\_est9.\*

10: gb\_est10.\*

11: gb\_est11.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	222	81.0	228	9	BZ352867 SALK_0827
2	152	55.5	152	9	BZ770600 SALK_1435
3	106.2	38.8	237	9	BH814587 SALK_0666
4	101.2	36.9	147	9	BZ352865 SALK_0827
5	63.2	23.1	691	9	BZ019485 oeg72a05
6	49	17.9	397	3	BH863458 BP863458
7	45.4	16.6	465	9	B52379 CIT-HSP-200
8	43.6	15.9	609	1	AI990512 ws40e09.x
9	43	15.7	321	8	R88734 YP93b07.sl
10	42.2	15.4	1204	10	CNS016E2
11	42	15.3	135	1	AI540689 tn72ali.x
12	42	15.3	289	1	AI873432 wf82f11.x
13	42	15.3	360	1	AI290306 qm02a04.x
14	42	15.3	418	1	AI620654 tu85c06.x
15	42	15.3	421	1	AI261719 qz31b01.x
16	42	15.3	448	1	AI357111 qx16g02.x
17	42	15.3	452	1	AI903013 ok45e06.s
18	42	15.3	462	1	AW014993 UI-H-B10p
19	42	15.3	477	1	AI151151 qc87f07.x
20	42	15.3	498	1	AW512290 xx72h07.x
21	42	15.3	553	5	BU689059 UI-CF-EC1
22	42	15.3	558	1	AW339459 xz91b03.x

23	42	15.3	610	3	BM995524
24	42	15.3	657	3	BQ015514
25	42	15.3	684	3	BQ003373
26	42	15.3	727	3	BQ006479
27	42	15.3	750	6	CA447735
28	42	15.3	2026	4	AF116653
29	41.6	15.2	283	3	BP423473
30	41.6	15.2	412	1	AI494002
31	41.6	15.2	418	1	AI708269
32	41.6	15.2	442	2	BE219714
33	41.6	15.2	553	1	AI797039
34	41.4	15.1	34	8	W96222
35	41.4	15.1	471	3	BM144644
36	41.4	15.1	473	3	BM150022
37	41.4	15.1	475	1	AA700613
38	41.4	15.1	481	1	AW051787
39	41.4	15.1	491	1	AI708670
40	41.4	15.1	534	2	BE856347
41	41.4	15.1	548	2	BE501745
42	40.4	14.7	392	1	AI284430
43	40	14.6	139	8	N46359
44	39.8	14.5	473	1	AA936680
45	39.8	14.5	528	1	AI435057

## ALIGNMENTS

### RESULT 1

BZ352867

LOCUS

DEFINITION

Arabidopsis thaliana genomic clone SALK\_082724.45.05.x, genomic survey sequence.

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

BZ352867 228 bp DNA linear GSS 14-NOV-2002  
SALK\_082724.45.05.x Arabidopsis thaliana TDNA insertion lines  
Arabidopsis thaliana genomic clone SALK\_082724.45.05.x, genomic survey sequence.

BZ352867 GI:24943729

GSS.

Arabidopsis thaliana (thale cress)

Arabidopsis thaliana

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons; rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsi

Alonso, J.M., Leisse, T.J., Barajas, P., Chen, H., Cheuk, R., Gadrinab, C., Jeske, A., Karnes, M., Kim, C.J., Parker, H., Prednis, L., Shinn, P., Zimmerman, J., and Ecker, J.R.

A sequence-indexed library of insertion mutations in the Arabidopsis Genome

Unpublished (2001)

Contact: Joseph R. Ecker

Salk Institute Genomic Analysis Laboratory (SIGNAL)

The Salk Institute for Biological Studies

10010 N. Torrey Pines Road, La Jolla, CA 92037, USA

Tel: 858 453 4100 x1752

Fax: 858 558 6379

Email: ecker@salk.edu

This is single pass sequence recovered from the left border of TDNA. This sequence lies within 300 bases of the 5' end of

AT3925882.

Class: TDNA tagged.

Location/Qualifiers

1. 228

/organism="Arabidopsis thaliana"

/mol\_type="genomic DNA"

/ecotype="Col-0"

/db\_xref="taxon:3702"

/clone="SALK\_082724.45.05.x"

/note="PCR was performed on Arabidopsis thaliana lines each of which contains one or more TDNA insertion elements. The resultant fragment for each line was directly sequenced to determine the genomic sequence at the site of insertion. Details of the protocols used can

```
be found at http://signal.salk.edu/tdna_protocols.html"

ORIGIN
Query Match      81.0%; Score 222; DB 9; Length 228;
Best Local Similarity 97.4%; Pred. No. 4.3e-46;
Matches 222; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 24 TTTTTCACACATGAGCTTCCTTTTATTCCAAAGTCATATAAAGTGTGAGGTGATGATCTACT 83
    |||||
Db 1 TTTTTCNCNCNTGGACTTCCTTTATTCNAAAGTCNATFANAGTGTGAGGTGATGATCTACT 60
    |||||

QY 84 TACGGTTTAAACATCGCATGATGTGTCATTAGCATCAATCTCCACCGTCCCAATTTTAT 143
    |||||
Db 61 TACGGTTTAAACATCGCATGATGTGTCATTAGCATCAATCTCCACCGTCCCAATTTAT 120
    |||||

QY 144 TAGTTGTTGACAAATATCGACCGTCTAAAGTTCACACCGCGCTATAAGAGTTTCATTAT 203
    |||||
Db 121 TAGTTGTTGACAAATATCGACCGTCTAAAGTTCACACCGCGCTATAAGAGTTTCATTAT 180
    |||||

QY 204 AAATTTTAGCAAAATAAATTCAGAAATAATTTTCTTGACTAAGCT 251
    |||||
Db 181 AAATTTTAGCAAAATAAATTCAGCAATAATTTTCTTGACTAAGCT 228
    |||||

RESULT 2
BZ770600      152 bp DNA linear GSS 13-MAR-2003
LOCUS
DEFINITION
SALK 143535.56.00.x Arabidopsis thaliana TDNA insertion lines
Arabidopsis thaliana genomic clone SALK_143535.56.00.x, genomic
survey sequence.
ACCESSION
BZ770600
VERSION
BZ770600.1 GI:28944284
KEYWORDS
GSS.
SOURCE
Arabidopsis thaliana (thale cress)
ORGANISM
Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE
1 (bases 1 to 152)
Alonso,J.M., Leisse,T.J., Barajas,P., Chen,H., Cheuk,R.,
Gadrinab,C., Jeske,A., Karnes,M., Kim,C.J., Parker,H., Prednis,L.,
Shinn,P., Zimmerman,J. and Ecker,J.R.
A Sequence-Indexed Library of Insertion Mutations in the
Arabidopsis Genome
Unpublished (2001)
CONTACT: Joseph R. Ecker
Salk Institute Genomic Analysis Laboratory (SIGAL)
The Salk Institute for Biological Studies
10010 N. Torrey Pines Road, La Jolla, CA 92037, USA
Tel: 858 453 4100 x1752
Fax: 858 558 6379
Email: ecker@salk.edu
This is single pass sequence recovered from the left border of
TDNA. This sequence lies within 300 bases of the 5' end of
At3G25882.
Class: TDNA tagged.
FEATURES
Location/Qualifiers
1..152
/organism="Arabidopsis thaliana"
/mol_type="genomic DNA"
/ecotype="Col-0"
/db_xref="taxon:3702"
/clone="SALK_143535.56.00.x"
/notes="PCR was performed on Arabidopsis thaliana lines
each of which contains one or more TDNA insertion
elements. The resultant fragment for each line was
directly sequenced to determine the genomic sequence at
the site of insertion. Details of the protocols used can
be found at http://signal.salk.edu/tdna_protocols.html"

ORIGIN
Query Match      55.5%; Score 152; DB 9; Length 152;
Best Local Similarity 97.4%; Pred. No. 1.4e-16;
Matches 111; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY 156 ATATGACCGTCTAAGTTCCACACCGCGCTATAAGAGTTTCATTATAAATTTTAGCAA 215
    |||||
Db 1 ATATGACCGTCTAAGTTCCACACCGCGCTATAAGAGTTTCATTATAAATTTTAGCAA 60
    |||||

QY 216 AATAAAATCAGCAATAAATTTTCTTGACTAAGCTTAAACGCGCGTTAATTTTC 274
    |||||
Db 61 AATAAAATCAGCAATAAATTTTCTTGACTAAGCTGAAGCGGCGGTAAACATTTCTC 119
    |||||
```

RESULT 4  
BZ352865  
LOCUS  
DEFINITION  
BZ352865 147 bp DNA linear GSS 14-NOV-2002  
Arabidopsis thaliana genomic clone SALK\_082722.32.45.x, genomic  
survey sequence.

ACCESSION  
BZ352865  
VERSION  
BZ352865.1 GI:24943727  
KEYWORDS  
GSS.  
SOURCE  
Arabidopsis thaliana (thale cress)  
ORGANISM  
Arabidopsis thaliana  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;  
rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsi.

REFERENCE  
1 (bases 1 to 147)  
Alonso, J.M., Leisse, T.J., Barajas, P., Chen, H., Cheuk, R.,  
Gadriab, C., Jeske, A., Karnes, M., Kim, C.J., Parker, H., Prednis, L.,  
Shinn, P., Zimmerman, J. and Ecker, J.R.  
A Sequence-Indexed Library of Insertion Mutations in the  
Arabidopsis Genome  
Unpublished (2001)  
Contact: Joseph R. Ecker  
Salk Institute Genomic Analysis Laboratory (SIGNAL)  
The Salk Institute for Biological Studies  
10010 N. Torrey Pines Road, La Jolla, CA 92037, USA  
Tel: 858 453 4100 x1752  
Fax: 858 558 6379  
Email: ecker@salk.edu  
This is single pass sequence recovered from the left border of  
TDNA. This sequence lies within 300 bases of the 5' end of  
AC3925882.  
Class: TDNA tagged.  
Location/Qualifiers  
1. .147  
/organism="Arabidopsis thaliana"  
/mol\_type="genomic DNA"  
/ecotype="Col-0"  
/db\_xref="taxon:3702"  
/clone="SALK\_082722.32.45.x"  
/notes="PCR was performed on Arabidopsis thaliana lines  
each of which contains one or more TDNA insertion  
elements. The resultant fragment for each line was  
directly sequenced to determine the genomic sequence at  
the site of insertion. Details of the protocols used can  
be found at [http://signal.salk.edu/tdna\\_protocols.html](http://signal.salk.edu/tdna_protocols.html)"

ORIGIN  
Query Match 36.9%; Score 101.2; DB 9; Length 147;  
Best Local Similarity 92.7%; Pred. No. 2.6e-15;  
Matches 139; Conservative 0; Mismatches 8; Indels 3; Gaps 3;  
QY 102 ATGATGATGTCATTAGCATCAATCTCCACCGCTCAATTTATTAGTTGTTGACAAATATCG 161  
DB 1 ATGATGATGTTAGTTAGTTGAATCTCCCTGTCCTCAATTT-TTTACTTTGTTGA-AAATATCG 58  
QY 162 ACCGTCTAAGTTCCACACCGCGCTATAGAGTTTCATTATATAATTTTAGCAAAATAAA 221  
DB 59 ACTGCTTAGTTCCAC-CCGACGCTATAGAGTTTCATTATATAATTTTAGCAAAATAAA 117  
QY 222 ATCAGCAATAATTTTCTTGTGATAAGCT 251  
DB 118 ATCAGCAATAATTTTCTTGTGATAAGCT 147

RESULT 5  
BZ019485  
LOCUS  
DEFINITION  
BZ019485 691 bp DNA linear GSS 08-OCT-2002  
Arabidopsis thaliana genomic clone SALK\_082722.32.45.x, genomic  
survey sequence.

ACCESSION  
BZ019485  
VERSION  
BZ019485.1 GI:23579218  
KEYWORDS  
GSS.  
SOURCE  
Brassica oleracea  
ORGANISM  
Brassica oleracea  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;  
rosids; eurosids II; Brassicales; Brassicaceae; Brassica.

REFERENCE  
1 (bases 1 to 691)  
Nash, W., Rabinowicz, P.D. and Wilson, R.K.  
Whole genome shotgun reads from Brassica oleracea  
Unpublished (2002)  
Contact: Richard K. Wilson  
Genome Sequencing Center  
Washington University School of Medicine  
Email: [submissions@watson.wustl.edu](mailto:submissions@watson.wustl.edu)  
Plate: oeg72 row: g column: 05  
Seq primer: -28RPOT reverse  
Class: shotgun  
High quality sequence start: 32  
High quality sequence stop: 551.  
Location/Qualifiers  
1. .691  
/organism="Brassica oleracea"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:3712"  
/clone lib="B.oleracea002"  
/note="Vector: pOTw13; Whole genome shotgun library from  
flowering buds. DNA was purified from a crude nuclear  
prep using Brassica oleracea 10100DH3 buds provided by  
Thomas Osborn at the University of Wisconsin. Genomic  
DNA was provided by Pablo Rabinowicz (CSHL) and the  
shotgun library prepared at Washington University Genome  
Sequencing Center."

ORIGIN  
Query Match 23.1%; Score 63.2; DB 9; Length 691;  
Best Local Similarity 67.7%; Pred. No. 1.4e-05;  
Matches 149; Conservative 0; Mismatches 63; Indels 8; Gaps 4;  
QY 6 AATATAGCGGATACCAATTTTCCACATGACGACTCTCTTTATTCGAAAGTCA-ATAA 64  
DB 465 AAGTAAGCAGATACCAAAATTTTC--CACATGACATCCCTTTATTCGAAAGTCCATTGA 522  
QY 65 AGTGTGACGTCATGATACCTTACGCTTTTAAACATCGCATGATGTCATTAGCATCAAT 124  
DB 523 AGCCACCATCCACTTTATTTGGCGTTGACGACATCGACGCTCAAAGTTCCACCGACG 578  
QY 125 CTCACCGTCCCAATTTTATTAGTTGTTGACAATATCGACCGTCTAAGTTCCACCGACG 184  
DB 579 CTCACCATCCACTTTATTTGGCGTTGACGACATCGACGCTCAAAGTTCCACCGACG 638  
QY 185 -GCTATAAGATTTCAATTATAAATTTTAGCAAAATAAAAT 223  
DB 639 TCCTATAAGAAGATGATCTCTCATATAAATCAAAATAAAAT 678

RESULT 6  
BP863458  
LOCUS  
DEFINITION  
BP863458 397 bp mRNA linear EST 17-FEB-2005  
Arabidopsis thaliana cdna clone RAFL21-66-M11 5',  
mRNA sequence.

ACCESSION  
BP863458  
VERSION  
BP863458.1 GI:59924431  
KEYWORDS  
EST.  
SOURCE  
Arabidopsis thaliana (thale cress)  
ORGANISM  
Arabidopsis thaliana  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;  
rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.

REFERENCE  
1 (bases 1 to 397)  
Seki, M., Ishida, J., Kamiya, A., Satou, M., Nakajima, M., Akiyama, K.,  
Iida, K., Enju, A., Sakurai, T., Arakawa, T., Carninci, P., Fukuda, S.,

VERSION  
BZ019485.1 GI:23579218  
KEYWORDS  
GSS.  
SOURCE  
Brassica oleracea  
ORGANISM  
Brassica oleracea  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;  
rosids; eurosids II; Brassicales; Brassicaceae; Brassica.

REFERENCE  
1 (bases 1 to 691)  
Nash, W., Rabinowicz, P.D. and Wilson, R.K.  
Whole genome shotgun reads from Brassica oleracea  
Unpublished (2002)  
Contact: Richard K. Wilson  
Genome Sequencing Center  
Washington University School of Medicine  
Email: [submissions@watson.wustl.edu](mailto:submissions@watson.wustl.edu)  
Plate: oeg72 row: g column: 05  
Seq primer: -28RPOT reverse  
Class: shotgun  
High quality sequence start: 32  
High quality sequence stop: 551.  
Location/Qualifiers  
1. .691  
/organism="Brassica oleracea"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:3712"  
/clone lib="B.oleracea002"  
/note="Vector: pOTw13; Whole genome shotgun library from  
flowering buds. DNA was purified from a crude nuclear  
prep using Brassica oleracea 10100DH3 buds provided by  
Thomas Osborn at the University of Wisconsin. Genomic  
DNA was provided by Pablo Rabinowicz (CSHL) and the  
shotgun library prepared at Washington University Genome  
Sequencing Center."

ORIGIN  
Query Match 23.1%; Score 63.2; DB 9; Length 691;  
Best Local Similarity 67.7%; Pred. No. 1.4e-05;  
Matches 149; Conservative 0; Mismatches 63; Indels 8; Gaps 4;  
QY 6 AATATAGCGGATACCAATTTTCCACATGACGACTCTCTTTATTCGAAAGTCA-ATAA 64  
DB 465 AAGTAAGCAGATACCAAAATTTTC--CACATGACATCCCTTTATTCGAAAGTCCATTGA 522  
QY 65 AGTGTGACGTCATGATACCTTACGCTTTTAAACATCGCATGATGTCATTAGCATCAAT 124  
DB 523 AGCCACCATCCACTTTATTTGGCGTTGACGACATCGACGCTCAAAGTTCCACCGACG 578  
QY 125 CTCACCGTCCCAATTTTATTAGTTGTTGACAATATCGACCGTCTAAGTTCCACCGACG 184  
DB 579 CTCACCATCCACTTTATTTGGCGTTGACGACATCGACGCTCAAAGTTCCACCGACG 638  
QY 185 -GCTATAAGATTTCAATTATAAATTTTAGCAAAATAAAAT 223  
DB 639 TCCTATAAGAAGATGATCTCTCATATAAATCAAAATAAAAT 678

RESULT 6  
BP863458  
LOCUS  
DEFINITION  
BP863458 397 bp mRNA linear EST 17-FEB-2005  
Arabidopsis thaliana cdna clone RAFL21-66-M11 5',  
mRNA sequence.

ACCESSION  
BP863458  
VERSION  
BP863458.1 GI:59924431  
KEYWORDS  
EST.  
SOURCE  
Arabidopsis thaliana (thale cress)  
ORGANISM  
Arabidopsis thaliana  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;  
rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.

REFERENCE  
1 (bases 1 to 397)  
Seki, M., Ishida, J., Kamiya, A., Satou, M., Nakajima, M., Akiyama, K.,  
Iida, K., Enju, A., Sakurai, T., Arakawa, T., Carninci, P., Fukuda, S.,

Iida,J., Kawai,J., Sasaki,D., Shiraki,T., Hayashizaki,Y. and Shinozaki,K.  
Large-Scale Analysis of RIKEN Arabidopsis Full-length cDNAs  
Unpublished (2005)  
Contact: Motoaki Seki  
Plant Functional Genomics Research Group  
RIKEN Genomic Sciences Center  
3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan  
Tel: 81-298-36-4359  
Fax: 81-298-36-9060  
Email: mseki@tc.riken.go.jp  
An Arabidopsis full-length cDNA library was constructed essentially as reported previously (Seki et al., 1998, 2002). This clone is in a modified plnuascript vector.  
Please visit our web site (<http://pfweb.gsc.riken.jp>) and <http://range.gsc.riken.jp> for further details.  
Location/Qualifiers  
1..397  
/organism="Arabidopsis thaliana"  
/mol\_type="mRNA"  
/db\_xref="taxon:3702"  
/clone="RAF21-66-M11"  
/lab\_host="DH10B"  
/clone\_lib="RAF21"  
/note="Site 1: BamHI; Site 2: SalI; Subtraction Library.  
The sequence was obtained from samples subjected to various stress and plant hormones-treated"

source  
FEATURES  
Query Match 17.9%; Score 49; DB 3; Length 397;  
Best Local Similarity 100.0%; Pred. NO. 0.056;  
Matches 49; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 226 GCAATAATTTTCTTGACTAGCTTAACGACGCCGTTAACTTTTC 274  
DB 1 GCAATAATTTTCTTGACTAGCTTAACGACGCCGTTAACTTTTC 49

RESULT 7  
B52379/c  
LOCUS B52379 465 bp DNA linear GSS 20-JUN-1998  
DEFINITION CIT-HSP-2003E3.TP CIT-HSP Homo sapiens genomic clone 2003E3,  
genomic survey sequence.  
ACCESSION B52379 1 GI:2606713  
VERSION B52379.1  
KEYWORDS GSS.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
Hominidae; Homo.  
REFERENCE 1 (bases 1 to 465)  
ADAMS,M.D., ROUNALEY,S.D., FIELD,C.E., BASS,S., LINHER,K.,  
GOLDEN,K., BERRY,K., GRANGER,D., SUH,E., WIBLE,C., SHIZUYA,H.,  
SIMON,M. and VENTER,J.C.  
Simon, M. and Venter, J.C.  
Use of a random BAC End Sequence Database for Sequence-Ready Map  
Building  
Unpublished (1997)  
Other-GSSs: CIT-HSP-2003E3.TR  
CONTACT: Mark Adams  
Department of Eukaryotic Genomics  
The Institute for Genomic Research  
9712 Medical Center Dr., Rockville, MD 20850, USA  
Tel: 301 838 0200  
Fax: 301 838 0208  
Email: mdadam@tigr.org  
Clones are available from Research Genetics ([info@resgen.com](mailto:info@resgen.com)). BAC  
end search page:  
[http://www.tigr.org/tdb/humgen/bac\\_end\\_search/bac\\_end\\_search.html](http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html)  
Seq primer: M33-21  
Class: BAC ends.  
Location/Qualifiers  
1..465  
source  
FEATURES  
Query Match 16.6%; Score 45.4; DB 9; Length 465;  
Best Local Similarity 56.3%; Pred. NO. 0.47; Indels 0; Gaps 0;  
Matches 85; Conservative 0; Mismatches 66; Indels 0; Gaps 0;  
QY 103 TGATGATGTCATTAGCATCAATCTCCACGCTCCAAATTTATTATTAGTTGTCACAAATATCA 162  
DB 388 TAATGAAGTAGGAGCAAGATTATCCATTGAAGGTATTATTATATGCAGCTCACTTAA 329  
QY 163 CGGTCTAAGTTCCACACCGAGCTATAAGAGTTTCATTATAAATTTTAGCAAAATATAAA 222  
DB 328 GTGACAAAATTCATACAGAGACTATAACAGAAATCATATTATAATATTAAAAATTAAT 269  
QY 223 TCAGCAAAATATTTTCTTGACTAAGCTTA 253  
DB 268 ACTTCAATATCTCTTTCACATTAAGATGATTA 238

ORIGIN  
Query Match 16.6%; Score 45.4; DB 9; Length 465;  
Best Local Similarity 56.3%; Pred. NO. 0.47; Indels 0; Gaps 0;  
Matches 85; Conservative 0; Mismatches 66; Indels 0; Gaps 0;  
QY 103 TGATGATGTCATTAGCATCAATCTCCACGCTCCAAATTTATTATTAGTTGTCACAAATATCA 162  
DB 388 TAATGAAGTAGGAGCAAGATTATCCATTGAAGGTATTATTATATGCAGCTCACTTAA 329  
QY 163 CGGTCTAAGTTCCACACCGAGCTATAAGAGTTTCATTATAAATTTTAGCAAAATATAAA 222  
DB 328 GTGACAAAATTCATACAGAGACTATAACAGAAATCATATTATAATATTAAAAATTAAT 269  
QY 223 TCAGCAAAATATTTTCTTGACTAAGCTTA 253  
DB 268 ACTTCAATATCTCTTTCACATTAAGATGATTA 238

RESULT 8  
A1990512  
LOCUS A1990512 609 bp mRNA linear EST 08-SEP-1999  
DEFINITION ws40e09.x1 NCI CGAP\_GC6 Homo sapiens cDNA clone IMAGE:2499688 3',  
mRNA sequence.  
ACCESSION A1990512  
VERSION A1990512.1 GI:5837393  
KEYWORDS EST.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
Hominidae; Homo.  
REFERENCE 1 (bases 1 to 609)  
NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
Tumor Gene Index  
Unpublished (1997)  
CONTACT: Robert Strausberg, Ph.D.  
Email: [cgapbs-r@mail.nih.gov](mailto:cgapbs-r@mail.nih.gov)  
Tissue Procurement: Christopher A. Moskaluk, M.D., Ph.D., Michael  
R. Emmert-Buck, M.D., Ph.D.  
cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima  
Bonaldo, Ph.D.  
cDNA Library Arrayed by: Greg Lennon, Ph.D.  
DNA Sequencing by: Washington University Genome Sequencing Center  
Clone distribution: NCI-CGAP clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
[www-bio.llnl.gov/bbrp/image/image.html](http://www-bio.llnl.gov/bbrp/image/image.html)  
Seq primer: -40UP from Gibco  
High quality sequence stop: 444.  
Location/Qualifiers  
1..609  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="IMAGE:2499688"  
/tissue\_type="pooled germ cell tumors"  
/lab\_host="DH10B"  
/clone\_lib="NCI CGAP GC6"  
/note="Vector: pT7T3D-Pac (Pharmacia) with a modified  
polylinker; Site 1: Not I; Site 2: Eco RI; Plasmid DNA  
from the normalized library NCI-CGAP GC4 was prepared, and  
ss circles were made in vitro. Following HAP purification,

```
ORIGIN
Query Match          15.4%; Score 42.2; DB 10; Length 1204;
Best Local Similarity 16.9%; Pred. No. 3.3;
Matches 43; Conservative 117; Mismatches 95; Indels 0; Gaps 0;

Qy      19   ACCAATTTTCCACACATGGCATTCCTTTATTCCAAGTCAATAAGTGAGCGTCATG 78
       :|::|||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::
Db      830  MMCMHTTTTTYTCCTMMGMHWNMTTHMHTTMTTMHMWNTTHTMTTWHTTTHMMHMMMTTT 889

       79   ATACTTAGCCTTTAAAACAATGCCGANGATGTGCATTAGCATCAATCTCCACCGTCCAAT 138
```









GenCore version 5.1.6  
Copyright (c) 1993 - 2006 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 11, 2006, 22:55:15 ; Search time 498 Seconds  
(without alignments)  
3666.919 Million cell updates/sec

Title: US-10-800-161-28

Perfect score: 274

Sequence: 1 tctagaatatagcgatcac.....acgacgcggttaacattttc 274

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 4996997 seqs, 3332346308 residues

Total number of hits satisfying chosen parameters: 9993994

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N Geneseq 21.\*

- 1: Geneseqn1980s.\*
- 2: Geneseqn1990s.\*
- 3: Geneseqn2000s.\*
- 4: Geneseqn2001as.\*
- 5: Geneseqn2001bs.\*
- 6: Geneseqn2002as.\*
- 7: Geneseqn2002bs.\*
- 8: Geneseqn2003as.\*
- 9: Geneseqn2003bs.\*
- 10: Geneseqn2003cs.\*
- 11: Geneseqn2003ds.\*
- 12: Geneseqn2004as.\*
- 13: Geneseqn2004bs.\*
- 14: Geneseqn2005s.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	274	100.0	274	13	Adri6847 Arabidops
2	272.4	99.4	274	13	Adri6845 Arabidops
3	272.4	99.4	544	13	Adri6846 Arabidops
4	272.4	99.4	862	13	Adri6844 Arabidops
5	272.4	99.4	862	14	Adv66225 Arabidops
6	272.4	99.4	962	13	Adri6843 Arabidops
7	272.4	99.4	1700	12	Adri6843 Arabidops
8	272.4	99.4	1700	13	Adri6822 Arabidops
9	272.4	99.4	1700	14	Adv66204 Arabidops
10	258.4	94.3	1226	14	Adw12661 A. thalia
11	52.8	19.3	509	12	Adri6820 Arabidops
12	52.8	19.3	509	13	Adri6820 Arabidops
13	52.8	19.3	509	14	Adv66202 Arabidops
14	43	15.7	321	6	Abz35565 Human gen
C 15	42	15.3	229	4	Aas25538 Human ova
C 16	42	15.3	1197	3	Aas52606 Eosinophi
C 17	42	15.3	2000	13	Adp5764 Human PRO
C 18	42	15.3	4641	8	Abz35897 Human sec
C 19	42	15.3	4673	6	Abz35431 Human gen

C 20	42	15.3	4779	10	ADF90715
C 21	41.4	15.1	2026	6	ABV77883
C 22	39.4	14.4	375	9	ACH50239
C 23	38.8	14.2	6301	6	ABL32053
C 24	38.2	13.9	2000	11	ACL35363
C 25	37.8	13.8	319608	3	AH51601
C 26	37.8	13.8	319608	5	AAS09301
C 27	37.2	13.6	1834	14	ADW26070
C 28	37.2	13.6	1834	14	ADW26105
C 29	37.2	13.6	1852	10	ADD22878
C 30	37.2	13.6	1963	10	ADD22877
C 31	37.2	13.6	1969	14	ADW26073
C 32	37.2	13.6	1969	14	ADW26106
C 33	37.2	13.6	2000	11	ACL37108
C 34	36	13.1	182328	12	ADL08128
C 35	35.6	13.0	87394	13	ADT55151
C 36	35.4	12.9	478	13	ACF91349
C 37	35.4	12.9	589	4	AAR80413
C 38	35.4	12.9	140036	6	AAS98600
C 39	35.2	12.8	621	8	ACA23093
C 40	35.2	12.8	110000	2	AAX20248_04
C 41	35	12.8	516	12	ACH68922
C 42	35	12.8	2000	11	ACL35887
C 43	34.8	12.7	2208	12	ADO35476
C 44	34.8	12.7	2208	13	ADS95227
C 45	34.8	12.7	2208	13	ADV68084

ALIGNMENTS

RESULT 1

ADRI6847

ID ADRI6847 standard; DNA; 274 BP.

XX AC ADRI6847;

XX DT 21-OCT-2004 (first entry)

XX DE Arabidopsis thaliana NI16 promoter DNA #5.

XX KW NI16; transgenic; herbicide resistance; mouse-ear cross; promoter; ds.

XX OS Arabidopsis thaliana.

XX PN US2004154051-A1.

XX PD 05-AUG-2004.

XX PF 12-MAR-2004; 2004US-00800161.

XX PR 15-DEC-1999; 99US-0171008P.

XX PR 11-JAN-2000; 2000US-0175519P.

XX PR 08-DEC-2000; 2000US-0073368S.

XX PR 20-JAN-2004; 2004US-00760752.

XX (SYGN ) SYNGENTA PARTICIPATIONS AG.

XX PA Cade RM, Dietrich RA, Lawton KA;

XX PI WPI; 2004-580223/56.

XX PT Novel isolated nucleic acid molecule comprising Arabidopsis NI16

XX FT promoter, useful in regulating transcription of coding sequence of

XX interest.

XX PS Claim 1; SEQ ID NO 28; 39pp; English.

XX CC The present invention relates to an Arabidopsis NI16 gene promoter useful

XX CC in regulating transcription of coding sequence of interest. The invention

XX CC is useful in production of transgenic plant or seed exhibiting resistance

XX CC to herbicides and microorganism such as Phytophthora parasitica, The

XX CC Pseudomonas syringae, Cercospora nicotianae, Peronospora parasitica. The



XX SQ Sequence 544 BP; 209 A; 84 C; 95 G; 156 T; 0 U; 0 Other;  
Query Match 99.4%; Score 272.4; DB 13; Length 544;  
Best Local Similarity 99.6%; Pred. No. 1.6e-63;  
Matches 273; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
QY 1 TCTAGAAATATAGCCGATACCAATTTTCCACACATGGACTTCCTTTATTTCCAAAAGTCA 60  
DB 271 TCTAAAAATATAGCCGATACCAATTTTCCACACATGGACTTCCTTTATTTCCAAAAGTCA 330  
QY 61 ATAAAGTGTGACGTCAATGATCTTACGCTTTTAAACATCGCATGATGTCATTAGCAT 120  
DB 331 ATAAAGTGTGACGTCAATGATCTTACGCTTTTAAACATCGCATGATGTCATTAGCAT 390  
QY 121 CAATCTCCACCGTCCAAATTTATTTAGTTGTGACAATATCGACCGTCTAAGTTCCACACC 180  
DB 391 CAATCTCCACCGTCCAAATTTATTTAGTTGTGACAATATCGACCGTCTAAGTTCCACACC 450  
QY 181 GACGGCTATAAGAGTTTCATTATAAATTTTAGCAAAATAAAATCAGCAAAATAATTTTTC 240  
DB 451 GACGGCTATAAGAGTTTCATTATAAATTTTAGCAAAATAAAATCAGCAAAATAATTTTTC 510  
QY 241 TTGACTAAGCTTAAAGACGCCGTTAAACATTTTC 274  
DB 511 TTGACTAAGCTTAAAGACGCCGTTAAACATTTTC 544

RESULT 4  
ADRI6844  
ID ADRI6844 standard; DNA; 862 BP.  
XX AC ADRI6844;  
XX DT 21-OCT-2004 (first entry)  
XX DE Arabidopsis thaliana N116 promoter DNA #2.  
XX KW N116; transgenic; herbicide resistance; mouse-ear cross; promoter; ds.  
XX OS Arabidopsis thaliana.

XX PN US2004154051-A1.  
XX PD 05-AUG-2004.  
XX PF 12-MAR-2004; 2004US-008000161.  
XX PR 15-DEC-1999; 99US-0171008P.  
XX PR 11-JAN-2000; 2000US-0175519P.  
XX PR 08-DEC-2000; 2000US-00733685.  
XX PR 20-JAN-2004; 2004US-00760752.  
XX PA (SYGN ) SYNGENTA PARTICIPATIONS AG.  
XX PI Cade RM, Dietrich RA, Lawton KA;  
XX WPI; 2004-580223/56.  
XX PT Novel isolated nucleic acid molecule comprising Arabidopsis N116  
XX promoter, useful in regulating transcription of coding sequence of  
XX interest.  
XX PS Claim 12; SEQ ID NO 25; 39pp; English.

XX CC The present invention relates to an Arabidopsis N116 gene promoter useful  
XX CC in regulating transcription of coding sequence of interest. The invention  
XX CC is useful in production of transgenic plant or seed exhibiting resistance  
XX CC to herbicides and microorganisms such as Phytophthora parasitica,  
XX CC pseudomonas syringae, Cercospora nicotianae, Peronospora parasitica. The  
XX CC present sequence is Arabidopsis thaliana N116 promoter DNA. This sequence  
XX CC is used in the exemplification of the invention.

SQ Sequence 862 BP; 337 A; 112 C; 152 G; 261 T; 0 U; 0 Other;  
Query Match 99.4%; Score 272.4; DB 13; Length 862;  
Best Local Similarity 99.6%; Pred. No. 1.8e-63;  
Matches 273; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
QY 1 TCTAGAAATATAGCCGATACCAATTTTCCACACATGGACTTCCTTTATTTCCAAAAGTCA 60  
DB 578 TCTAAAAATATAGCCGATACCAATTTTCCACACATGGACTTCCTTTATTTCCAAAAGTCA 637  
QY 61 ATAAAGTGTGACGTCAATGATCTTACGCTTTTAAACATCGCATGATGTCATTAGCAT 120  
DB 638 ATAAAGTGTGACGTCAATGATCTTACGCTTTTAAACATCGCATGATGTCATTAGCAT 697  
QY 121 CAATCTCCACCGTCCAAATTTATTTAGTTGTGACAATATCGACCGTCTAAGTTCCACACC 180  
DB 698 CAATCTCCACCGTCCAAATTTATTTAGTTGTGACAATATCGACCGTCTAAGTTCCACACC 757  
QY 181 GACGGCTATAAGAGTTTCATTATAAATTTTAGCAAAATAAAATCAGCAAAATAATTTTTC 240  
DB 758 GACGGCTATAAGAGTTTCATTATAAATTTTAGCAAAATAAAATCAGCAAAATAATTTTTC 817  
QY 241 TTGACTAAGCTTAAAGACGCCGTTAAACATTTTC 274  
DB 818 TTGACTAAGCTTAAAGACGCCGTTAAACATTTTC 851

RESULT 5  
ADV66225  
ID ADV66225 standard; DNA; 862 BP.  
XX AC ADV66225;  
XX DT 24-FEB-2005 (first entry)  
XX DE Arabidopsis thaliana N116 promoter.  
XX KW Systemic acquired resistance; disease-resistance; transgenic plant;  
XX KW plant; N116; N1M1 interactor; promoter; ds.  
XX OS Arabidopsis thaliana.  
XX Key Location/Qualifiers  
XX FH promotor 1..862  
XX FT /\*tag= a  
XX FT /note= "This promoter is specifically claimed"  
XX FT misc\_feature 365..374  
XX FT /\*tag= b  
XX FT /note= "TCAI motif"  
XX FT misc\_feature 426..435  
XX FT /\*tag= c  
XX FT /note= "TCAI motif"  
XX FT enhancer 609..614  
XX FT /\*tag= d  
XX FT /note= "MYCATR22 element"  
XX FT enhancer 646..665  
XX FT /\*tag= e  
XX FT /note= "CMV AS1 salicylic acid response element"  
XX FT enhancer 707..712  
XX FT /\*tag= f  
XX FT /note= "PAL BOX"  
XX FT enhancer 757..762  
XX FT /\*tag= g  
XX FT /note= "HEXAMERAT 4 element"  
XX PN US2004248303-A1.  
XX PD 09-DEC-2004.  
XX PF 20-JAN-2004; 2004US-00760752.  
XX PR 15-DEC-1999; 99US-0171008P.  
XX PR 11-JAN-2000; 2000US-0175519P.

```
PR 08-DEC-2000; 2000US-00733685.
XX (SYGN ) SYNGENTA PARTICIPATIONS AG.
PA Cade RM, Dietrich RA;
PI WPI; 2005-020569/02.
XX P-PSDB; ADV66203.
XX Novel isolated N116 nucleic acid promoter, useful for initiating
PT transcription of DNA that encodes proteins involved in regulation of
PT systemic acquired resistance (SAR) gene expression in plants.
XX Claim 1; Page; 32pp; English.
XX The invention relates to a N116 nucleic acid promoter sequence. The
CC promoter sequence of the invention is useful for initiating transcription
CC of DNA that encodes proteins involved in regulation of systemic acquired
CC resistance (SAR) gene expression in plants. It enhances the expression of
CC SAR genes such as pathogenesis-related protein (PR-1), thus allows broad-
CC spectrum disease resistance in plant such as resistance against
CC Pseudomonas syringae. The present sequence is the Arabidopsis thaliana
CC N116 gene including 5' upstream promoter. Note: This sequence is not
CC given separately, but has been created from the N116 gene (ADV66204).
CC Given in the sequence listing and the information provided in claim 1.
XX
SQ Sequence 862 BP; 337 A; 112 C; 152 G; 261 T; 0 U; 0 Other;
Query Match 99.4%; Score 272.4; DB 14; Length 862;
Best Local Similarity 99.6%; Pred. No. 1.8e-63;
Matches 273; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 TCTGAATATATAGCCGATACCAATTTTCCACACATGGACTTCCTTTATTCCAAAGTCA 60
DB 578 TCTAAATATAGCCGATACCAATTTTCCACACATGGACTTCCTTTATTCCAAAGTCA 637
QY 61 ATAAAGTGTGACGTATGATCTAGCTTTTAAACATCGATGATGTCATTAGCAT 120
DB 638 ATAAAGTGTGACGTATGATCTAGCTTTTAAACATCGATGATGTCATTAGCAT 697
QY 121 CAATCTCCACCGTCCCAATTTATTTAGTGTGACAAATATCGACCGTCTAAGTTCCACACC 180
DB 698 CAATCTCCACCGTCCCAATTTATTTAGTGTGACAAATATCGACCGTCTAAGTTCCACACC 757
QY 181 GACGGCTATAAGAGTTTCATTATAAATTTTAGCAAAATAAAATCAGCAAAATAATTTTTC 240
DB 758 GACGGCTATAAGAGTTTCATTATAAATTTTAGCAAAATAAAATCAGCAAAATAATTTTTC 817
QY 241 TTGACTAAGCTTAAACGACCGCGTTAAACAATTTTC 274
DB 818 TTGACTAAGCTTAAACGACCGCGTTAAACAATTTTC 851
RESULT 6
ADRI6843
ID ADRI6843 standard; DNA; 962 BP.
XX
XX AC ADRI6843;
XX
XX DT 21-OCT-2004 (first entry)
XX
XX DE Arabidopsis thaliana N116 promoter DNA #1.
XX
XX KW N116; transgenic; herbicide resistance; mouse-ear cress; promoter; ds.
XX
XX OS Arabidopsis thaliana.
XX
XX PN US2004154051-A1.
XX
XX PD 05-AUG-2004.
XX
XX PF 12-MAR-2004; 2004US-00800161.
```

```
PR 15-DEC-1999; 99US-0171008P.
PR 11-JAN-2000; 2000US-0175519P.
PR 08-DEC-2000; 2000US-00733685.
PR 20-JAN-2004; 2004US-00760752.
XX (SYGN ) SYNGENTA PARTICIPATIONS AG.
XX Cade RM, Dietrich RA, Lawton KA;
XX WPI; 2004-580223/56.
XX Novel isolated nucleic acid molecule comprising Arabidopsis N116
PT promoter, useful in regulating transcription of coding sequence of
PT interest.
XX Claim 11; SEQ ID NO 24; 39pp; English.
XX The present invention relates to an Arabidopsis N116 gene promoter useful
CC in regulating transcription of coding sequence of interest. The invention
CC is useful in production of transgenic plant or seed exhibiting resistance
CC to herbicides and microorganism such as Phytophthora parasitica,
CC Pseudomonas syringae, Cercospora nicotianae, Peronospora parasitica. The
CC present sequence is Arabidopsis thaliana N116 promoter DNA. This sequence
CC is used in the exemplification of the invention.
XX
SQ Sequence 962 BP; 379 A; 125 C; 165 G; 293 T; 0 U; 0 Other;
Query Match 99.4%; Score 272.4; DB 13; Length 962;
Best Local Similarity 99.6%; Pred. No. 1.8e-63;
Matches 273; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 TCTGAATATATAGCCGATACCAATTTTCCACACATGGACTTCCTTTATTCCAAAGTCA 60
DB 678 TCTAAATATATAGCCGATACCAATTTTCCACACATGGACTTCCTTTATTCCAAAGTCA 737
QY 61 ATAAAGTGTGACGTATGATCTAGCTTTTAAACATCGATGATGTCATTAGCAT 120
DB 738 ATAAAGTGTGACGTATGATCTAGCTTTTAAACATCGATGATGTCATTAGCAT 797
QY 121 CAATCTCCACCGTCCCAATTTATTTAGTGTGACAAATATCGACCGTCTAAGTTCCACACC 180
DB 798 CAATCTCCACCGTCCCAATTTATTTAGTGTGACAAATATCGACCGTCTAAGTTCCACACC 857
QY 181 GACGGCTATAAGAGTTTCATTATAAATTTTAGCAAAATAAAATCAGCAAAATAATTTTTC 240
DB 858 GACGGCTATAAGAGTTTCATTATAAATTTTAGCAAAATAAAATCAGCAAAATAATTTTTC 917
QY 241 TTGACTAAGCTTAAACGACCGCGTTAAACAATTTTC 274
DB 918 TTGACTAAGCTTAAACGACCGCGTTAAACAATTTTC 951
RESULT 7
ADNI7259
ID ADNI7259 standard; DNA; 1700 BP.
XX
XX AC ADNI7259;
XX
XX DT 17-JUN-2004 (first entry)
XX
XX DE Arabidopsis thaliana N116 gene.
XX
XX KW N116; SAR; systemic acquired resistance; mouse-ear cress; gene; ds.
XX
XX OS Arabidopsis thaliana.
XX
XX FH Key Location/Qualifiers
FT misc_feature 365..374
FT /*tag= a
FT /note= "TCAL motif"
FT misc_feature 426..435
FT /*tag= b
FT /note= "TCAL motif"
```

```
FT misc_feature 609..614 /*tag= c
FT /*note= "MYCATR22 element"
FT misc_feature 646..665 /*tag= d
FT /*note= "CAMV AS1 salicylic acid response element"
FT misc_feature 707..712 /*tag= e
FT /*note= "PAL box"
FT misc_feature 757..762 /*tag= f
FT /*note= "HEXAMERAT 4 element"
FT CDS 863..1231 /*tag= g
FT /*product= "Arabidopsis thaliana N116 protein"
XX
PN US6706952-B1.
XX
XX 16-MAR-2004.
XX
XX 08-DEC-2000; 2000US-00733685.
XX
XX 15-DEC-1999; 99US-0171008P.
XX 11-JAN-2000; 2000US-0175519P.
XX
XX (SYGN ) SYNGENTA PARTICIPATIONS AG.
XX
XX Cad RM, Dietrich RA;
XX
XX WPI; 2004-313378/29.
XX P-PSDB; ADN17258.
XX
XX New N116 nucleic acid sequence and encoded protein, useful for increasing
XX systemic acquired resistance gene expression in a plant.
XX
XX Example 2; SEQ ID NO 3; 29pp; English.
XX
XX The invention relates to Arabidopsis N116 gene encoding a protein
XX involved in the regulation of SAR gene expression in plants. The N116
XX nucleic acid molecule and the encoded protein is useful in increasing
XX systemic acquired resistance (SAR) gene expression in a plant. The
XX present sequence is Arabidopsis thaliana N116 gene.
XX
XX Sequence 1700 BP; 593 A; 237 C; 361 G; 509 T; 0 U; 0 Other;
XX
XX Query Match 99.4%; Score 272.4; DB 12; Length 1700;
XX Best Local Similarity 99.6%; Pred. No. 2e-63;
XX Matches 273; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Qy 1 TCTAGAAATATAGCGGATACCAATTTTCCACATGGACTTCCTTTATTCCAAAGTCA 60
Db |||||
578 TCTAAAAATATAGCGGATACCAATTTTCCACATGGACTTCCTTTATTCCAAAGTCA 637
Qy |||||
61 ATAAAGTGTGACGTCATGATCTTACGCTTTAAACATCGCATGATGTCATTAGCAT 120
Db |||||
638 ATAAAGTGTGACGTCATGATCTTACGCTTTAAACATCGCATGATGTCATTAGCAT 697
Qy |||||
121 CAATCTCACCGTCCAAATTTATTTAGTTGTTCACAATATCGACGCTCTAAGTTCACACC 180
Db |||||
698 CAATCTCACCGTCCAAATTTATTTAGTTGTTCACAATATCGACGCTCTAAGTTCACACC 757
Qy |||||
181 GACGCTATAAGAGTTTCATTATATAATTTTAGCAAAATAAATCAGCAAAATATTTTTC 240
Db |||||
758 GACGCTATAAGAGTTTCATTATATAATTTTAGCAAAATAAATCAGCAAAATATTTTTC 817
Qy |||||
241 TTGACTAGCTTAACGACGCGGTTTAAACATTTTC 274
Db |||||
818 TTGACTAGCTTAACGACGCGGTTTAAACATTTTC 851
XX
XX RESULT 8
XX ADR16822
XX ID ADR16822 standard; DNA; 1700 BP.
```

```
XX ADR16822;
XX AC
XX 21-OCT-2004 (first entry)
XX
XX Arabidopsis thaliana N116 genomic DNA.
XX
XX N116; transgenic; herbicide resistance; mouse-ear cross; ds; gene.
XX Arabidopsis thaliana.
XX
XX Key Location/Qualifiers
XX misc_feature 365..374 /*tag= a
XX /*note= "TCA1 motif"
XX misc_feature 426..435 /*tag= b
XX /*note= "TCA1 motif"
XX misc_feature 609..614 /*tag= c
XX /*note= "MYCATR22 element"
XX misc_feature 646..665 /*tag= d
XX /*note= "CAMV AS1 salicylic acid response element"
XX misc_feature 707..712 /*tag= e
XX /*note= "PAL BOX"
XX misc_feature 757..762 /*tag= f
XX /*note= "HEXAMERAT 4 element"
XX CDS 863..1231 /*tag= g
XX /*product= "Arabidopsis thaliana N116 protein"
XX
XX US2004154051-A1.
XX
XX 05-AUG-2004.
XX
XX 12-MAR-2004; 2004US-00800161.
XX
XX 15-DEC-1999; 99US-0171008P.
XX 11-JAN-2000; 2000US-0175519P.
XX 08-DEC-2000; 2000US-00733685.
XX 20-JAN-2004; 2004US-00760752.
XX
XX (SYGN ) SYNGENTA PARTICIPATIONS AG.
XX
XX Cade RM, Dietrich RA, Lawton KA;
XX
XX WPI; 2004-580223/56.
XX P-PSDB; ADR16821.
XX
XX Novel isolated nucleic acid molecule comprising Arabidopsis N116
XX promoter, useful in regulating transcription of coding sequence of
XX interest.
XX
XX Example 3; SEQ ID NO 3; 39pp; English.
XX
XX The present invention relates to an Arabidopsis N116 gene promoter useful
XX in regulating transcription of coding sequence of interest. The invention
XX is useful in production of transgenic plant or seed exhibiting resistance
XX to herbicides and microorganism such as Phytophthora parasitica,
XX Pseudomonas syringae, Cercospora nicotianae, Peronospora parasitica. The
XX present sequence is Arabidopsis thaliana N116 genomic DNA, including the
XX 5'upstream promoter sequence.
XX
XX Sequence 1700 BP; 593 A; 237 C; 361 G; 509 T; 0 U; 0 Other;
XX
XX Query Match 99.4%; Score 272.4; DB 13; Length 1700;
XX Best Local Similarity 99.6%; Pred. No. 2e-63;
XX Matches 273; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Qy 1 TCTAGAAATATAGCGGATACCAATTTTCCACATGGACTTCCTTTATTCCAAAGTCA 60
```



PI Pfizner A, Roth B;  
 XX WPI; 2005-114423/13.  
 XX Novel nucleic acid containing selectively inducible NIMIN-1 promoter  
 PT sequence or NIMIN-2 promoter sequence, useful in controlled expression of  
 PT desired expression products in suitable host expression organisms such as  
 PT transgenic plants.  
 XX Claim 1; SEQ ID NO 2; 22pp; English.  
 XX The invention relates to a novel nucleic acid (I) containing at least a  
 CC first nucleotide sequence chosen from a NIMIN-1 promoter sequence of  
 CC (ADW12660) and NIMIN-2 promoter sequence of (ADW12661), and its  
 CC biologically active derivative. The polynucleotide of the invention is  
 CC useful in controlled expression of desired expression products in  
 CC suitable host expression organisms such as transgenic plants. The present  
 CC sequence represents the NIMIN-2 promoter.  
 XX Query Match 94.3%; Score 258.4; DB 14; Length 1226;  
 XX Best Local Similarity 99.6%; Pred. No. 1.1e-59;  
 XX Matches 259; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 QY 1 TCTAGAAATATAGCGATACCAATTTTCCACACATGGAGCTTCCTTTATTTCCAAAGTCA 60  
 Db 967 TCTAAATATATAGCGATACCAATTTTCCACACATGGAGCTTCCTTTATTTCCAAAGTCA 1026  
 QY 61 ATAAAGTGTGACGTGATGATCTTAAGCTTTTAAACATCGCATGATGTCATTAGCAT 120  
 Db 1027 ATAAAGTGTGACGTGATGATCTTAAGCTTTTAAACATCGCATGATGTCATTAGCAT 1086  
 QY 121 CAATCTCCACCGTCCAAATTTATTAGTTGTGTGACAATATCGACGCTTAAGTTCACACC 180  
 Db 1087 CAATCTCCACCGTCCAAATTTATTAGTTGTGTGACAATATCGACGCTTAAGTTCACACC 1146  
 QY 181 GACGGCTATAGAGTTTCATTAATAATTTTTCGAAATTAATTAATTAATTAATTTTTC 240  
 Db 1147 GACGGCTATAGAGTTTCATTAATAATTTTTCGAAATTAATTAATTAATTTTTC 1206  
 QY 241 TTGACTAAGCTTAAACGACG 260  
 Db 1207 TTGACTAAGCTTAAACGACG 1226  
 RESULT 11  
 ADN17257  
 ID ADN17257 standard; cDNA; 509 BP.  
 AC ADN17257;  
 XX 17-JUN-2004 (first entry)  
 DE Arabidopsis thaliana NI16 cDNA.  
 XX NI16; SAR; systemic acquired resistance; mouse-ear cross; gene; ss.  
 XX Arabidopsis thaliana.  
 XX Key Location/Qualifiers  
 FT CDS 68..436  
 FT /\*tag= a  
 FT /product= "Arabidopsis thaliana NI16 protein"  
 XX US6706952-B1.  
 XX 16-MAR-2004.  
 XX 08-DEC-2000; 2000US-00733685.  
 XX 15-DEC-1999; 99US-0171008P.  
 XX 11-JAN-2000; 2000US-0175519P.

XX (SYGN ) SYNGENTA PARTICIPATIONS AG.  
 XX Cad RM, Dietrich RA;  
 XX WPI; 2004-313378/29.  
 XX P-PSDB; ADN17258.  
 XX New NI16 nucleic acid sequence and encoded protein, useful for increasing  
 PT systemic acquired resistance gene expression in a plant.  
 XX Claim 2; SEQ ID NO 1; 29pp; English.  
 XX The invention relates to Arabidopsis NI16 gene encoding a protein  
 CC involved in the regulation of SAR gene expression in plants. The NI16  
 CC nucleic acid molecule and the encoded protein is useful in increasing  
 CC systemic acquired resistance (SAR) gene expression in a plant. The  
 CC present sequence is Arabidopsis thaliana NI16 cDNA.  
 XX Query Match 19.3%; Score 52.8; DB 12; Length 509;  
 XX Best Local Similarity 96.4%; Pred. No. 0.00021;  
 XX Matches 54; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
 QY 219 AAAATCAGCAATAAATTTTCTTGACTAAGCTTAAACGACGCGTAAACATTTTC 274  
 Db 1 AAAATCAGCAATAAATTTTCTTGACTAAGCTTAAACGACGCGTAAACATTTTC 56  
 RESULT 12  
 ADN16820  
 ID ADN16820 standard; cDNA; 509 BP.  
 AC ADN16820;  
 XX 21-OCT-2004 (first entry)  
 DE Arabidopsis thaliana NI16 cDNA.  
 XX NI16; transgenic; herbicide resistance; mouse-ear cross; gene.  
 XX Arabidopsis thaliana.  
 XX Key Location/Qualifiers  
 FT CDS 68..436  
 FT /\*tag= a  
 FT /product= "Arabidopsis thaliana NI16 protein"  
 XX US2004154051-A1.  
 XX 05-AUG-2004.  
 XX 12-MAR-2004; 2004US-00800161.  
 XX 15-DEC-1999; 99US-0171008P.  
 XX 11-JAN-2000; 2000US-0175519P.  
 XX 08-DEC-2000; 2000US-00733685.  
 XX 20-JAN-2004; 2004US-00760752.  
 XX (SYGN ) SYNGENTA PARTICIPATIONS AG.  
 XX Cade RM, Dietrich RA, Lawton KA;  
 XX WPI; 2004-580223/56.  
 XX P-PSDB; ADN16821.  
 XX Novel isolated nucleic acid molecule comprising Arabidopsis NI16  
 PT promoter, useful in regulating transcription of coding sequence of  
 PT interest.  
 XX Example 1; SEQ ID NO 1; 39pp; English.

CC The present invention relates to an Arabidopsis N116 gene promoter useful  
 CC in regulating transcription of coding sequence of interest. The invention  
 CC is useful in production of transgenic plant or seed exhibiting resistance  
 CC to herbicides and microorganism such as Phytophthora parasitica,  
 CC Pseudomonas syringae, Cercospora nicotianae, Peronospora parasitica. The  
 CC present sequence is Arabidopsis thaliana N116 cDNA.  
 XX  
 SQ Sequence 509 BP; 158 A; 82 C; 133 G; 136 T; 0 U; 0 Other;

Query Match 19.3%; Score 52.8; DB 13; Length 509;  
 Best Local Similarity 96.4%; Pred. No. 0.00021;  
 Matches 54; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 219 AAAATCAGCAATAATTTTCTTGACTAAAGCTTAAACGACCGCTTAACATTTC 274  
 |||||||  
 DB 1 AAAATCAGCAATAATTTTCTTGACTAAAGCTTAAACGACCGCTTAACATTTC 56

RESULT 13  
 ADV66202  
 ID ADV66202 standard; cDNA; 509 BP.  
 AC ADV66202;  
 XX  
 DT 24-FEB-2005 (first entry)  
 XX  
 DE Arabidopsis thaliana N116 cDNA.  
 KW Systemic acquired resistance; disease-resistance; transgenic plant;  
 KW plant; N116; N1M1 interactor; gene; ss.  
 XX Arabidopsis thaliana.

Key Location/Qualifiers  
 CDS 68..436  
 FT /\*tag= a  
 FT /product= "Arabidopsis thaliana N116 protein"  
 XX

US2004248303-A1.  
 PD 09-DEC-2004.  
 XX  
 PF 20-JAN-2004; 2004US-00760752.  
 XX  
 PR 15-DEC-1999; 99US-0171008P.  
 PR 11-JAN-2000; 2000US-0175519P.  
 PR 08-DEC-2000; 2000US-00733685.  
 XX (SYGN ) SYNGENTA PARTICIPATIONS AG.  
 PA Cade RM, Dietrich RA;  
 PI WPI: 2005-020569/02.  
 DR P-PSDB; ADV66203.  
 XX  
 PT Novel isolated N116 nucleic acid promoter, useful for initiating  
 PT transcription of DNA that encodes proteins involved in regulation of  
 PT systemic acquired resistance (SAR) gene expression in plants.  
 XX  
 PS Example 1; SEQ ID NO 1; 32pp; English.

CC The invention relates to a N116 nucleic acid promoter sequence. The  
 CC promoter sequence of the invention is useful for initiating transcription  
 CC of DNA that encodes proteins involved in regulation of systemic acquired  
 CC resistance (SAR) gene expression in plants. It enhances the expression of  
 CC SAR genes such as pathogenesis-related protein (PR-1), thus allows broad-  
 CC spectrum disease resistance in plant such as resistance against  
 CC Pseudomonas syringae. The present sequence is the Arabidopsis thaliana  
 CC N116 cDNA obtained via a N1M1 two-hybrid screen.  
 XX

SQ Sequence 509 BP; 158 A; 82 C; 133 G; 136 T; 0 U; 0 Other;

Query Match 19.3%; Score 52.8; DB 14; Length 509;

Best Local Similarity 96.4%; Pred. No. 0.00021;  
 Matches 54; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
 QY 219 AAAATCAGCAATAATTTTCTTGACTAAAGCTTAAACGACCGCTTAACATTTC 274  
 |||||||  
 DB 1 AAAATCAGCAATAATTTTCTTGACTAAAGCTTAAACGACCGCTTAACATTTC 56

RESULT 14  
 ABZ35565  
 ID ABZ35565 standard; cDNA; 321 BP.  
 XX  
 AC ABZ35565;  
 XX  
 DT 05-FEB-2003 (first entry)  
 XX  
 DE Human gene expression profile polynucleotide SEQ ID NO 676.

Human; artery; endothelium; umbilical; vein; aorta; pulmonary artery;  
 KW bronchial epithelium; prostate; muscle; lung fibroblast; osteoblast;  
 KW tumour; microarray; genome mapping; antibiotic; antiviral; antifungal;  
 KW gene expression; gene; ss.

OS Homo sapiens.

XX WO200274979-A2.

PN 26-SEP-2002.

PD 20-MAR-2002; 2002WO-US008456.

PF 20-MAR-2001; 2001US-0276947P.

XX (ORTH ) ORTHO CLINICAL DIAGNOSTICS INC.

PA Wan J, Wang Y;

PI WPI: 2002-740862/80.

DR New gene expression profile generated from primary, endothelial,  
 PT epithelial, and muscle cell types, useful for identifying disease  
 PT pathologies involving alterations of gene expression, e.g. cancer.

XX Example 3; Page 799; 850pp; English.

CC The invention relates to a gene expression profile comprising one or more  
 CC genes (ABZ34889-ABZ35692) and generated from a cell type. The cell type  
 CC is a coronary artery endothelium, umbilical artery or vein endothelium,  
 CC aortic endothelium, dermal microvascular endothelium, pulmonary artery  
 CC endothelium, myometrium microvascular endothelium, keratinocyte  
 CC epithelium, bronchial epithelium, mammary epithelium, prostate  
 CC epithelium, renal cortical epithelium, renal epithelium, umbilical artery smooth  
 CC muscle, neonatal dermal fibroblast, pulmonary artery smooth muscle,  
 CC dermal fibroblast, neural progenitor cells, skeletal muscle, astrocytes,  
 CC aortic smooth muscle, mesangial cells, coronary artery smooth muscle,  
 CC bronchial smooth muscle, uterine smooth muscle, lung fibroblast,  
 CC osteoblasts or prostate stromal cell. The gene expression profile is used  
 CC for determining the level of RNA expression for a sample, determining the  
 CC phenotype of a cell and distinguishing cell types. The gene or a protein  
 CC expression profile is useful in identifying disease pathologies involving  
 CC alterations of gene expression. The assessment of expression profiles may  
 CC provide meaningful information with respect to tumour type and stage,  
 CC treatment methods, and prognosis. The gene or protein expression profile  
 CC may also be used for creating microarrays. The microarray is useful for  
 CC genetic and physical mapping of genomes, DNA sequencing, genetic or  
 CC medical diagnosis, genotyping of organisms, confirming cell or tissue  
 CC identifications and in identifying promising antibiotics, antiviral or  
 CC antifungal agents

SQ Sequence 321 BP; 112 A; 51 C; 40 G; 113 T; 0 U; 5 Other;

Query Match 15.7%; Score 43; DB 6; Length 321;





***This Page Blank (uspto)***